DOCUMENT RESUME

ED 120 228 TM 005 187

AUTHOR Noe, Michael J.
TITLE A Monte Carlo Study of Several Test Procedures in the Repeated Measures Design.
PUB DATE Apr 76
NOTE 66p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage
DESCRIPTORS *Analysis of Variance; *Comparative Analysis; Computer Programs; Correlation; *Mathematical Models; Matrices; Simulation; Statistical Analysis; *Tests of Significance
IDENTIFIERS Monte Carlo Method; *Repeated Measures; Type I Error

ABSTRACT

This study compared three approaches to the two-factor experiment with repeated measures on one factor: (1) the conventional mixed model analysis of variance, (2) the Greenhouse-Geisser conservative analysis of variance, and (3) multivariate extensions of analysis of variance. Computer simulated data were used in a total of 96 sets of covariance structures and sample size ratio to determine type I error rates when the usual covariance matrix restrictions of compound symmetry and equality across levels of the nonrepeated factor are violated jointly and separately. Practical implications of the results are discussed. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED120228

Session 28.11

A MONTE CARLO STUDY OF SEVERAL TEST PROCEDURES IN THE REPEATED MEASURES DESIGN

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Michael J. Noe

University of Illinois College of Medicine

Paper presented at the 1976 Annual Meeting of the American Educational Research Association, San Francisco, California, April 19-23, 1976.

TM005 187

ABSTRACT

A Monte Carlo Study of Several Test Procedures in the Repeated Measures Design

MICHAEL J. NOE, University of Illinois College of Medicine

This study compared three approaches to the two-factor experiment with repeated measures on one factor: (1) the conventional mixed model analysis of variance, (2) the Greenhouse-Geisser conservative analysis of variance, and (3) multivariate extensions of analysis of variance. Computer simulated data were used in a total of 96 sets of covariance structures and sample size ratios to determine type I error rates when the usual covariance matrix restrictions of compound symmetry and equality across levels of the non-repeated factor are violated jointly and separately. Practical implications of the results are discussed.

CHAPTER 1

PRESENTATION OF THE THREE APPROACHES CONSIDERED

In his review of the history of repeated measures, Collier (1958) noted that concern over experiments involving repeated measurements was manifested quite early in statistics and the social sciences. Perhaps the first analysis of designs in which observations were assumed correlated (repeated measurements is only one such design in this broader class) was made by Yates (1933) who presented the analysis of the split-plot design as it related to agricultural experimentation. Until about fifteen years ago, most of the work relative to the split-plot design had been done by individuals outside of the fields of education and psychology. If there are no carry-over effects, the split-plot design is analogous to a repeated measures design in the area of behavioral sciences; where additive carry-over effects are present, repeated measures designs are analogous to crossover designs (Winer, 1971, Ch. 7).

Advantages of the repeated measures design include control over individual differences between experimental units and economy of subjects. Disadvantages include the damaging effects of fatigue, practice, transfer of learning, and the effects of an immediately preceding success or failure. Randomization, rather than remove them, only completely confounds these effects with treatment effects and is therefore useful only if the sequence effects are likely to be minimal relative to treatment effects. Randomization is not possible to use when the factor to be studied is the effect of time or position, as is the case in learning experiments.

Repeated measures data can take many forms. It may take the form of developmental data where measures are obtained on each subject at successive points in time. In learning data the repeated measures represent scores on successive blocks of response trials. In situations where the subject serves as his own control, the resulting repeated measures data correspond to the responses of the subject to the several treatments which may or may not have been presented to him in random order.

For the sake of uniformity, the repeated measures dimension, hereafter called factor B, will be referred to as "occasions." The levels of the non-repeated factor A used in this paper will be denoted as "treatments"; that is, mutually exclusive groups of subjects observed under I different treatment conditions. It should be noted, though, that the levels of factor A could just as well represent sub-populations or varieties of subjects.

1.1 Theoretical Model for All Three Procedures

The development of the following layout to be used throughout this paper parallels that of Collier (1958) and Danford, Hughes, and McNee (1960):

Let there be I levels of some treatment with n_i subjects exposed to the i^{th} level, and each subject measured on some characteristic for K occasions following treatment. Let X_{ijk} be the measurement made at the k^{th} occasion on the j^{th} individual at the i^{th} treatment level. Then with each subject for a given i^{th} treatment level there is a vector of K measurements

$$\underset{\sim}{x}'_{j(i)} = [X_{ij1}, X_{ij2}, \ldots, X_{ijK}]$$

where $i = 1,2,\ldots,I$ and $j = 1,2,\ldots,n_i$.

The average over the population from which the subjects are drawn is denoted by

$$E(X_{ijk}) = \mu_{ik} = \mu + \alpha_i + \beta_k + \gamma_{ik}$$

where the fixed effects constants are defined as

μ = overall mean,

α_i = effect of i^{th} treatment level,

β_k = k^{th} occasion effect,

and

γ_{ik} = treatment x occasion interaction effect,

with the conditions that

$$\sum_i n_i\alpha_i = 0, \quad \sum_k \beta_k = 0, \text{ and } \sum_i n_i\gamma_{ik} = \sum_k \gamma_{ik} = 0 .$$

The random departure of X_{ijk} from μ_{ik} is expressed as m_{ijk} so that

$$X_{ijk} = \mu_{ik} + m_{ijk} ,$$

where m_{ijk} is decomposed into the random effects $d_{j(i)} + g_{j(i)k}$, or into $d_{j(i)} + f_{j(i)k} + e_{ijk}$ with the symbols defined as

$d_{j(i)}$ = effect of the j^{th} individual measured at the i^{th} treatment level,

$g_{j(i)k}$ = individual x occasion interaction effect, $f_{j(i)k}$, plus random error component, e_{ijk}.

The random components m_{ijk} are assumed to follow a K-variate normal distribution with

$$E(m_{ijk}) = 0$$

and a covariance matrix with elements

$$\mathrm{Cov}(m_{ijk}, m_{i'j'k'}) = \mathrm{Cov}(X_{ijk}, X_{i'j'k'}) = \begin{cases} \sigma_{kk}, & \text{if } i = i',\ j = j',\ k = k' \\ \sigma_{kk'}, & \text{if } i = i',\ j = j',\ k \neq k' \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, $\underset{\sim}{x}'_{j(i)}$ is a random vector sampled from a K-variate normal distribution with mean vector

$$\underset{\sim}{\mu}'_{(i)} = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{iK}],$$

and an arbitrary covariance matrix

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & \vdots & & \vdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{bmatrix},$$

where $\underset{\sim}{\Sigma}$ is assumed to be the same for the various treatment groups (i.e., for all levels of I).

The model equation which is appropriate for both our univariate and multivariate procedures can be written as

$$X_{ijk} = \mu + \alpha_i + d_{j(i)} + \beta_k + \gamma_{ik} + g_{j(i)k},$$

where $i = 1,2,\ldots,I$; $j = 1,2,\ldots,n_i$; $k = 1,2,\ldots,K$.

1.2 Conventional Mixed Model ANOVA Procedure

When $K > 2$ (as will be assumed throughout this paper), a further assumption is needed for the case of the conventional mixed model ANOVA procedure in order that the usually computed ratios of mean squares be

exactly distributed as the F-distribution -- it is necessary for the K variates to have equal variances and be mutually independent or, at most, have equal correlations. Thus, the covariance matrix $\underset{\sim}{\Sigma}$ must have the form

$$\underset{\sim}{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

where ρ is the common correlation coefficient.

This highly restrictive assumption of compound symmetry is rarely satisfied in most behavioral investigations. For example, among repeated measures taken in growth and change studies, the measurements taken closer together in time are often more highly correlated than those taken further apart. The matrix is then said to have the so-called simplex form and serial correlation is said to be present.

The summary table for the conventional mixed model ANOVA procedure, the first of three repeated measures approaches to be considered, is presented in Table 1.

1.3 Greenhouse-Geisser Conservative ANOVA Procedure

Box (1954b) investigated the effects of inequality of variance and first order serial correlation of errors in the two-way classification ANOVA (N rows and K columns). He found that when the appropriate null hypothesis was true, inequality of variance from column to column increased the level of significance (probability of Type I error) for the test on homogeneity of column means, but that first order serial correlation

Table 1

Two Factor ANOVA with Repeated Measures on Factor B
(both factors assumed fixed)

Source	d.f.	Sum of Squares	$\mathcal{F}$
B (occasion)	K-1	$Q_1 = N\sum_k(\bar{X}_{..k}-\bar{X}_{...})^2$	$\mathcal{F}_B = \frac{Q_1}{Q_5/(N-I)}$
A (treatment)	I-1	$Q_2 = K\sum_i n_i(\bar{X}_{i..}-\bar{X}_{...})^2$	$\mathcal{F}_A = \frac{Q_2/(I-1)}{Q_3/(N-I)}$
Subjects within treatment	N-I	$Q_3 = K\sum\sum_{ij}(\bar{X}_{ij.}-\bar{X}_{i..})^2$	
A × B	(I-1)(K-1)	$Q_4 = \sum\sum_{ik} n_i(\bar{X}_{i.k}-\bar{X}_{..k}-\bar{X}_{i..}+\bar{X}_{...})^2$	$\mathcal{F}_{AB} = \frac{Q_4/(I-1)}{Q_5/(N-I)}$
B × subjects within treatment	(N-I)(K-1)	$Q_5 = \sum\sum\sum_{ijk}(X_{ijk}-\bar{X}_{i.k}-\bar{X}_{ij.}+\bar{X}_{i..})^2$	
Total	NK-1	$Q_6 = \sum\sum\sum_{ijk}(X_{ijk}-\bar{X}_{...})^2$	

(Note: The usual dot notation applies; e.g., $\bar{X}_{ij.}$ is the mean of the j^{th} subject in the i^{th} treatment, $\bar{X}_{i..}$ is the i^{th} treatment mean, and $\bar{X}_{...}$ is the overall mean.)

within rows had little effect on the $\mathcal{F}$-test. Using a formula developed earlier (Box, 1954a) for the approximation to the distribution of the ratio of two independent quadratic forms, Box showed that the null distribution of $\mathcal{F}$ was approximately

$$\mathcal{F}[(K-1)\epsilon,(K-1)(N-1)\epsilon] ,$$

where

$$\epsilon = K^2(\bar{\sigma}_{kk}-\bar{\sigma}_{..})^2/[(K-1)\cdot(\Sigma\Sigma_{kk'}\sigma^2_{kk'}-2K\Sigma_{k}\bar{\sigma}^2_{k.}+K^2\bar{\sigma}^2_{..})],$$

$\sigma_{kk'}$ is an element of the matrix $\underset{\sim}{\Sigma}$,

$\bar{\sigma}_{kk}$ is the mean of the diagonal elements of $\underset{\sim}{\Sigma}$,

$\bar{\sigma}_{k.}$ is the mean of the k^{th} row (or k^{th} column) of $\underset{\sim}{\Sigma}$,

and

$\bar{\sigma}_{..}$ is the mean of all elements of $\underset{\sim}{\Sigma}$.

The work of Box was actually in reference to one group of N experimental units. Geisser and Greenhouse (1958), and in a subsequent article (Greenhouse and Geisser, 1959), extended his work to I groups and found that $\mathcal{F}_B$ in Table 1 was approximately distributed like

$$\mathcal{F}[(K-1)\epsilon,(K-1)(N-I)\epsilon]$$

and $\mathcal{F}_{AB}$ was approximately distributed like

$$\mathcal{F}[(I-1)(K-1)\epsilon,(K-1)(N-I)\epsilon],$$

while $\mathcal{F}_A$ was exactly distributed like

$$\mathcal{F}[(I-1),(N-I)] .$$

They further showed that

$$1/(K-1) \leq \epsilon \leq 1 .$$

If the matrix $\underset{\sim}{\Sigma}$ has compound symmetry, as is assumed under the conventional ANOVA procedure, then ϵ equals unity and there is no reduction in the degrees of freedom for the $\mathcal{F}_B$ and $\mathcal{F}_{AB}$ tests. Taking $\epsilon = 1/(K-1)$, the lower bound, Greenhouse and Geisser provided "conservative" hypothesis tests for $\mathcal{F}_B$ and $\mathcal{F}_{AB}$ ($\mathcal{F}_A$ is unaffected by the composition of $\underset{\sim}{\Sigma}$)as follows:

$\mathcal{F}_B$ is conservatively distributed as $\mathcal{F}[1,(N-I)]$.

$\mathcal{F}_{AB}$ is conservatively distributed as $\mathcal{F}[(I-1),(N-I)]$.

These tests are called conservative because the minimum value of ϵ, $1/(K-1)$, gives rise to larger $\mathcal{F}$-values. No assumption on the composition of $\underset{\sim}{\Sigma}$ is made for these two tests. The $\mathcal{F}_A$ test for A-effects is not conservatively distributed because it results from a one-way ANOVA with all observations having the same variance ($\underset{\sim}{\Sigma}$ is assumed the same for all levels of factor A).

1.4 Multivariate Procedure

The third approach for repeated measures to be considered is through multivariate extensions of analysis of variance. As in the Greenhouse-Geisser conservative ANOVA procedure, the covariance matrix $\underset{\sim}{\Sigma}$ is permitted to be of arbitrary composition; however, as in the two previous procedures, $\underset{\sim}{\Sigma}$ must be identical for all levels of the non-repeated factor A.

1.4.1 History of Multivariate Test Criteria

Hotelling (1931) found the distribution of a quantity T^2 which was a multivariate extension of Student's t-distribution. Wilks (1932) used the likelihood ratio principle of Neyman and Pearson to develop a statistic Λ which constituted a multivariate extension of the $\mathcal{F}$-test in ANOVA. Thus, Hotelling's T^2 statistic is applicable to two-group

multivariate problems, and Wilks' Λ is an extension to more than two groups. Rulon and Brooks (1968) and Tatsuoka (1971, Ch. 6) showed that algebraically Wilks' Λ and Hotelling's T^2 are functionally related in the two-group case.

The characteristic roots of $\underset{\sim}{B}\underset{\sim}{W}^{-1}$ may be denoted as $\lambda_1, \lambda_2, \ldots, \lambda_p$, where $\underset{\sim}{W}$ is the error sums of products and cross products matrix and $\underset{\sim}{B}$ is the sums of products and cross products matrix for the effect being tested. In this context, Wilks' Λ is equal to the inverse of the product of the characteristic roots of $(\underset{\sim}{I} + \underset{\sim}{B}\underset{\sim}{W}^{-1})$, where $\underset{\sim}{I}$ is the identity matrix of the same order as $\underset{\sim}{B}$ and $\underset{\sim}{W}$; that is,

$$\Lambda = 1/[(1+\lambda_1)(1+\lambda_2)\ldots(1+\lambda_p)].$$

Hotelling (1951) proposed the distribution of the trace of the product matrix $\underset{\sim}{B}\underset{\sim}{W}^{-1}$. This statistic, called Hotelling's trace criterion τ, is equal to the sum of the characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\underset{\sim}{B}\underset{\sim}{W}^{-1}$ and may be shown as

$$\tau = \sum_{i=1}^{p} (\underset{\sim}{B}\underset{\sim}{W}^{-1})_{ii} = \sum_{i=1}^{p} \lambda_i .$$

A third approach by Roy (1957, Ch. 6) involved the distribution of the largest characteristic root of $\underset{\sim}{B}(\underset{\sim}{W} + \underset{\sim}{B})^{-1}$. All three of these tests are equivalent when there is one degree of freedom for effect being tested. A Monte Carlo analysis by Schatzoff (1966b) used a criterion called expected significance level (a measure somewhat similar to power) and found that Hotelling's trace criterion τ and Wilks' Λ produced better results against a wider spectrum of alternatives than Roy's largest root, but that with large samples the three statistics were

asymptotically similar. However, no clear cut choice between Hotelling's trace criterion τ and Wilks' Λ could be made, and consequently both were recommended.

Bartlett (1947) showed that a logarithmic function of Λ was distributed approximately as a chi-square, and Schatzoff (1966a) derived exact distributions for the statistic when either the number of variables or the number of degrees of freedom for the effect being tested was even. Rao (1952) gave an $\mathcal{F}$-approximation for another function of Λ which reduced to an exact test when either the number of variables or the degrees of freedom for the effect being tested equaled 1 or 2.

1.4.2 Multivariate Test for A-Effects

The test for A-effects in the theoretical model of Section 1.1 is not a multivariate problem. The test can be reduced to a univariate $\mathcal{F}$-test (assuming no interaction) as follows (see Morrison, 1967, Ch. 4):

Let the hypothesis for equal treatment levels be

$$\mathcal{H}_o: \underset{\sim}{p}'\underset{\sim}{\mu}_{(1)} = \underset{\sim}{p}'\underset{\sim}{\mu}_{(2)} = \cdots = \underset{\sim}{p}'\underset{\sim}{\mu}_{(I)} ,$$

where $\underset{\sim}{p}' = [1,1,\ldots,1]$ is the K-component vector with unity in each position, and $\underset{\sim}{\mu}'_{(i)} = [\mu_{i1},\mu_{i2},\ldots,\mu_{iK}]$ is the population mean vector of $\underset{\sim}{x}'_{j(i)}$ under the i^{th} treatment level. The $\mathcal{F}_A$ statistic in Table 1 is used to test $\mathcal{H}_o$. When $\mathcal{H}_o$ is true, $\mathcal{F}_A$ is distributed as $\mathcal{F}[(I-1),(N-I)]$.

1.4.3 Multivariate Test for B-Effects

The multivariate test for B-effects (occasion effects in our terminology) may be reduced to a problem involving Hotelling's T^2 statistic. Rao (1952, Ch. 7) first suggested the use of T^2 to test the equality of the K-1 contrasts involving the K variates. This approach was further elaborated on by Anderson (1958, Ch. 5) whose method is

modified for the I-treatment case as follows (assuming no interaction):

Let us test the hypothesis

$$\mathcal{H}_o: \beta_1 = \beta_2 = \ldots = \beta_K = 0$$

on the basis of a sample of size N of $\underset{\sim}{x}_{j(i)}$ from $\eta(\underset{\sim}{\mu}_{(i)}, \underset{\sim}{\Sigma})$. Let $\underset{\sim}{C}$ be any (K-1) x K matrix of rank K-1 such that

$$\underset{\sim\sim}{C}\underset{\sim}{p} = \underset{\sim}{0} ,$$

where $\underset{\sim}{p}' = [1,1,\ldots,1]$ and $\underset{\sim}{0}' = [0,0,\ldots,0]$. Each row of $\underset{\sim}{C}$ must sum to zero and the K-1 row contrasts must be linearly independent of one another. If

$$\underset{\sim}{y}_{j(i)} = \underset{\sim\sim}{C}\underset{\sim}{x}_{j(i)} \qquad (i = 1,2,\ldots,I;\ j = 1,2,\ldots,n_i)$$

then $\underset{\sim}{y}_{j(i)}$ has mean vector $\underset{\sim\sim}{C}\underset{\sim}{\mu}_{(i)}$ and covariance matrix $\underset{\sim\sim\sim}{C\Sigma C'}$. Let the vector of variate means in the population over all treatments be

$$\underset{\sim}{\mu}' = [\mu_1, \mu_2, \ldots, \mu_K] .$$

Then the hypothesis $\mathcal{H}_o$ is $\underset{\sim\sim}{C}\underset{\sim}{\mu} = \underset{\sim}{0}$. The statistic to test $\mathcal{H}_o$ is Hotelling's single sample $\mathcal{T}^2$ statistic, where

$$\mathcal{T}^2 = N\underset{\sim}{\bar{y}}'\underset{\sim}{D}^{-1}\underset{\sim}{\bar{y}} = N\underset{\sim}{\bar{x}}'\underset{\sim}{C}'(\underset{\sim\sim\sim}{CSC'})^{-1}\underset{\sim\sim}{C}\underset{\sim}{\bar{x}} ,$$

$$\underset{\sim}{\bar{x}} = \frac{1}{N}\sum_{i}\sum_{j} \underset{\sim}{x}_{j(i)} ,$$

$$\underset{\sim}{\bar{y}} = \frac{1}{N}\sum_{i}\sum_{j} \underset{\sim}{y}_{j(i)} = \underset{\sim\sim}{C}\underset{\sim}{\bar{x}} ,$$

$\underset{\sim}{D}$ is the within-treatments covariance matrix for the $\underset{\sim}{y}_{j(i)}$, and

$\underset{\sim}{S}$ is the within-treatments covariance matrix for the $\underset{\sim}{x}_{j(i)}$.

When the null hypothesis is true, the quantity

$$\mathcal{T}_B = \frac{N-I-K+2}{(K-1)(N-I)}\,\mathcal{J}^2$$

is distributed as $\mathcal{F}[(K-1),(N-I-K+2)]$.

1.4.4 Multivariate Tests for A × B-Effects

In order to use the multivariate test for treatment × occasion interaction effects in the general case of I treatments, one must compute the elements of the product matrix $\underset{\sim\sim}{B}\underset{\sim\sim}{W}^{-1}$ referred to in Section 1.4.1. In the present context, $\underset{\sim}{B}$ and $\underset{\sim}{W}$ are defined as the between-treatments and within-treatments sums of squares and cross products matrices respectively of the $\underset{\sim}{y}_{j(i)}$, not the $\underset{\sim}{x}_{j(i)}$.

As an example, suppose $\underset{\sim}{C}$ in Section 1.4.3 is chosen so that

$$\underset{\sim}{C} = \frac{1}{K}\begin{bmatrix} K-1 & -1 & -1 & \cdots & -1 & -1 \\ -1 & K-1 & -1 & \cdots & -1 & -1 \\ \vdots & & & & & \vdots \\ -1 & -1 & -1 & \cdots & K-1 & -1 \end{bmatrix}.$$

Then

$$\underset{\sim}{y}_{j(i)} = \underset{\sim\sim}{C}\underset{\sim}{x}_{j(i)} = \begin{bmatrix} X_{ij1} - \bar{X}_{ij\cdot} \\ X_{ij2} - \bar{X}_{ij\cdot} \\ \vdots \\ X_{ij(K-1)} - \bar{X}_{ij\cdot} \end{bmatrix} = \begin{bmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ij(K-1)} \end{bmatrix}.$$

Thus, the mean of each individual, $\bar{X}_{ij\cdot}$, is subtracted in turn from each of the first K-1 variates, so that

$$\underset{\sim}{y}'_{(i)} = [\bar{Y}_{i\cdot 1}, \bar{Y}_{i\cdot 2}, \ldots, \bar{Y}_{i\cdot (K-1)}]$$

is multivariate normal with mean

$$\underset{\sim}{\eta}'_{(i)} = [\eta_{i1}, \eta_{i2}, \ldots, \eta_{i(K-1)}], \quad \eta_{ik} = \mu_{ik} - \mu_{k} ,$$

and covariance matrix $\underset{\sim\sim}{C\Sigma C'}$. The null hypothesis is then stated as

$$\mathcal{H}_o : \underset{\sim}{\eta}_{(1)} = \underset{\sim}{\eta}_{(2)} \cdots = \underset{\sim}{\eta}_{(I)} .$$

The between-treatments sums of squares and cross products matrix $\underset{\sim}{B} = [b_{kk'}]$ has the element $b_{kk'}$ as

$$b_{kk'} = \sum_{i} n_i (\bar{Y}_{i\cdot k} - \bar{Y}_{\cdot\cdot k})(\bar{Y}_{i\cdot k'} - \bar{Y}_{\cdot\cdot k'}) ,$$

where $k,k' = 1,2,\ldots,K-1$.

The within-treatments sums of squares and cross products matrix $\underset{\sim}{W} = [w_{kk'}]$ has the element $w_{kk'}$ as

$$w_{kk'} = \sum_{i}\sum_{j} (Y_{ijk} - \bar{Y}_{i\cdot k})(Y_{ijk'} - \bar{Y}_{i\cdot k'}) .$$

Once the product matrix $\underset{\sim\sim}{BW}^{-1}$ is computed, one may resort to any of the multivariate test criteria of Section 1.4.1 (Hotelling's trace, Wilks' Λ, or Roy's largest root) to test the significance of the treatment x occasion interaction effect.

CHAPTER 2

FOCUS OF THE PRESENT STUDY

2.1 Statement of the Problem

The repeated measures model used in this Monte Carlo study was the two-factor experiment in which there were repeated measures on factor B, and both factors were considered fixed. The levels of factor A typically represent treatment levels or classification levels for groups of subjects, and the experimenter is concerned with the question of the statistical significance of the main effects for factors A and B, and of their interaction effect.

Three approaches to the solution of this problem - the conventional mixed model ANOVA, the Greenhouse-Geisser conservative ANOVA, and the multivariate procedures of Section 1.2-1.4 were compared in regard to the effect upon the probability of Type I error (level of significance) of violating one or both of the following assumptions:

1. that the population covariance matrix of repeated measures is the same for each of the levels of factor A (i.e., $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \ldots = \underset{\sim}{\Sigma}_I$ - required of all three approaches), and
2. that upon combining all levels of A, the repeated measures covariance matrix $\underset{\sim}{\Sigma}$ has equal variances and equal covariances in the population (this assumption of compound symmetry is required only of the mixed model ANOVA).

The question for which answers were sought may be stated as follows: "What are the actual probabilities of Type I error connected with each of the three different test procedures in relation to testing the significance of factors A, B, and A × B when the preceding homogeneity assumptions are violated in a specified manner?"

2.2 Relation to Previous Studies

The basic configuration utilized in this study for the model common to all three approaches was the two-factor setup with three levels of the non-repeated factor A, four levels of repeated factor B, and a total of 15 subjects over all levels of factor A; namely a configuration in which $I = 3$, $K = 4$, and $N = 15$. This was one of the two basic configurations considered by Collier et al. (1967), except that in their study equal numbers of subjects were chosen for each level of factor A. The present study considered situations in which there existed unequal numbers of subjects for the treatment levels.

The choices for the population covariance matrix patterns (i.e., composition of $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, $\underset{\sim}{\Sigma}_3$, and $\underset{\sim}{\Sigma}$) were based upon specific configurations employed by Collier et al. (1967) and Ito and Schull (1964). Eight patterns of treatment covariance matrices (and accompanying sample size ratios) used by Ito and Schull were considered in combination with twelve population covariance patterns for $\underset{\sim}{\Sigma}$ used by Collier et al., resulting in a total of 96 (8·12) sets of covariances and sample size ratios.

Ito and Schull were interested in the effect on the level of significance, with Hotelling's trace criterion τ as the measure, of violating the assumption that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_3$; whereas Collier et al. were concerned with the effect on level of significance, using the univariate mixed model ANOVA and Greenhouse-Geisser ANOVA procedures as measures, of violating the assumption that $\underset{\sim}{\Sigma}$ had compound symmetry. The present study goes further in concerning itself with combinations of covariance patterns used by the two previous studies, so that a comparison of one multivariate and two univariate repeated measures

procedures may be made when the two homogeneity assumptions of Section 2.1 are violated simultaneously.

2.3 Experimental Procedure

Empirical estimates of level of significance were made at the four following nominal α-levels: .10, .05, .025, and .01. These significance level estimates for an effect were estimates of the probabilities that the test statistics for the three approaches exceeded the four upper end cut-off points as defined by their null distributions. For example, the three significance level estimates for the tests for B-effects (occasion effects) were

$$\mathcal{P}\{\mathcal{F}_B > \mathcal{F}_{1-\alpha}[(K-1),(K-1)(N-I)] \,|\, \beta_i = 0\} ,$$

$$\mathcal{P}\{\mathcal{F}_B > \mathcal{F}_{1-\alpha}[1,(N-I)] \,|\, \beta_i = 0\} ,$$

and

$$\mathcal{P}\{\mathcal{J}_B > \mathcal{F}_{1-\alpha}[(K-1),(N-I-K+2)] \,|\, \beta_i = 0\} ,$$

for the conventional mixed model ANOVA, conservative ANOVA, and multivariate procedures, respectively.

The multivariate test for A x B-effects (treatment x occasion interaction effects) was Rao's $\mathcal{F}$-approximation statistic (Rao, 1952) which reduced to an exact $\mathcal{F}$ variate in this study because the hypothesis degrees of freedom equals I-1 = 2. Rao's statistic, as it related here to the A x B-effects test, is defined as

$$R = \frac{1 - \Lambda^{1/2}}{\Lambda^{1/2}} \cdot \frac{N-K-1}{K-1} ,$$

where $\Lambda = \dfrac{|\underset{\sim}{W}|}{|\underset{\sim}{B}+\underset{\sim}{W}|}$, with $\underset{\sim}{B}$ and $\underset{\sim}{W}$ the hypothesis and error sums of products

and cross products matrices, respectively. Under the null hypothesis, the $\mathcal{R}$ statistic follows the $\mathfrak{F}[2(K-1),2(N-K-1)]$ distribution exactly.

As noted in Section 1.4.2, the mixed model ANOVA and multivariate procedures for testing A-effects (treatment effects) lead to identical results. Therefore, estimates of level of significance for A-effects were confined to the conventional mixed model ANOVA procedure (there is no conservative ANOVA test for factor A).

2.3.1 Monte Carlo Simulation

The empirical investigation of these test statistics could properly be described as a Monte Carlo simulation. Cooley and Lohnes (1968, p. 87) defined a Monte Carlo study as an empirical study of statistics based upon random numbers. Various devices and techniques exist for producing sequences of random numbers. The simplest example is for a discrete probability distribution function where the choice is between N different numbers, each occurring with equal probability. A roulette wheel, each of whose sections corresponds to a different value, or a die with that number of faces will generate such a random sequence. It is from this connection with gambling that the name Monte Carlo has become a general term used to describe any computational method using random numbers.

Monte Carlo methods are used in a wide variety of problems such as the evaluation of complicated integrals, studies of population growth, and the design of nuclear reactors. Gordon (1969, p. 91) classified such cases as Monte Carlo simulations. These simulations require sequences of random numbers which are drawn from a distribution that, in general, is not uniform. The present study required random numbers drawn from a multivariate normal distribution. Methods for generating

random numbers from a particular distribution are usually not available. However, methods exist for generating uniformly distributed random numbers and then transforming that sequence into the required sequence.

2.3.2 Choices for Sample Sizes and Population Covariance Patterns

Fixing the sample size and number of variates deprives one of the opportunity of making inferences based upon empirical results as to their effect upon the level of significance. Collier et al. (1967) obtained similar estimates for level of significance when the number of subjects per level of factor A was increased from 5 to 15. The number of variates may exert a greater influence on the significance level as was shown by the results of Ito and Schull (1964) which suggested that effects on the level of significance may tend to increase as the number of variates increases. Answers to the questions of influence of sample size and number of variates on the significance level, deserving of investigation though they may be, were considered outside the scope of the present study.

The choices for population covariance matrix patterns and sample sizes for levels of factor A were arbitrary. Since specific configurations were required, the Monte Carlo simulation was conducted on the 96 (8·3·4) combinations of treatment sample sizes and covariance matrices presented in Tables 2-4. A given set was determined by selecting from one of the eight sets of Table 2, one of the three sets of Table 3, and one of the four sets of Table 4.

The 96 combinations were assigned unique three-digit numbers by letting the first digit denote the appropriate set in Table 2, the second digit denote the variance pattern in Table 3, and the third integer represent the correlation pattern in Table 4. For example,

Table 2

Population Treatment Covariance Matrices*
and Accompanying Sample Sizes

	Population Treatment Covariance Matrices			Number of Observations Sampled from Treatment Levels			Total Sample Size
	Σ_1	Σ_2	Σ_3				
Set	a_1	a_2	a_3	n_1	n_2	n_3	N
1	1	1	1	5	5	5	15
2	1	2	3	5	5	5	15
3	1	1	3	5	5	5	15
4	1	2	3	7	5	3	15
5	1	1	3	7	5	3	15
6	1	1	3	9	5	1	15
7	1	2	3	3	5	7	15
8	1	1	3	1	5	9	15

*The population treatment covariance matrix Σ_i ($i = 2,3$) is a_i times Σ_1.

Table 3

Variance Pattern of $\underset{\sim}{\Sigma}^*$

	Variance, σ_{ij}			
Set	j: 1	2	3	4
1	100	100	100	100
2	100	400	900	1600
3	100	225	400	900

Table 4

Correlation Pattern of $\underset{\sim}{\Sigma}^*$

	Correlation, $\rho_{jj'}$					
Set	jj': 12	13	14	23	24	34
1	.20	.20	.20	.20	.20	.20
2	.80	.80	.80	.80	.80	.80
3	.50	.25	.00	.25	.00	.00
4	.80	.50	.20	.80	.50	.80

*$\underset{\sim}{\Sigma}$ is the population covariance matrix over all treatments.

the set 821 had population treatment covariance matrices $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$ whose elements were in the numerical ratios of 1:1:3, respectively. Samples of size 15 from these matrices were computer simulated with 1, 5, and 9 observations sampled from $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$, respectively. The population covariance matrix over all treatments, $\underset{\sim}{\Sigma}$, had diagonal elements $\sigma_{11} = 100$, $\sigma_{22} = 400$, $\sigma_{33} = 900$, $\sigma_{44} = 1600$, and off-diagonal elements whose intercorrelations all equaled .20.

The population treatment covariance matrices in Table 2 were selected from among those presented by Ito and Schull (1964). The number of observations sampled from each treatment level within a given set were in the same numerical ratios as those of Ito and Schull (their N tended to infinity). The sets in Tables 3 and 4 were taken from among those used by Collier et al. (1967). In so doing, it was possible to have a common reference from which to draw conclusions based on empirical evidence.

2.3.3 Computer Simulation of Multivariate Normal Data

In order to obtain the empirical results, the random vector $\underset{\sim}{x}'_{j(i)} = [X_{ij1}, X_{ij2}, \ldots, X_{ijK}]$ was generated by means of the University of Illinois IBM 360/75 computer. As is detailed in Section 1.1, the components X_{ijk} of $\underset{\sim}{x}_{j(i)}$ are expressible as the sum of the fixed effects μ_{ik} and the random effects m_{ijk}. The vector $\underset{\sim}{x}_{j(i)}$ was assumed for purposes of this study to be a random vector sampled from a K-variate normal distribution with the covariance matrix $\underset{\sim}{\Sigma}_i$ and the mean vector $\underset{\sim}{\mu}'_{(i)} = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{iK}]$.

The matrix $\underset{\sim}{\Sigma}_i$ was factored into $\underset{\sim}{L}\underset{\sim}{L}'$, where $\underset{\sim}{L}$ was a lower triangular matrix. This triangular decomposition of the symmetric, positive semi-definite matrix $\underset{\sim}{\Sigma}_i$ is known by various names such as Cholesky

factorization and the square root method. An outline of the computing procedures involved in triangular decomposition was given by Harman (1967, Ch. 3) and also by Naylor et al. (1966, Ch. 4).

The computer generation of a given $\underset{\sim}{x}_{j(i)}$ first required the generation of an independent random vector $\underset{\sim}{z}' = [Z_1, Z_2, \ldots, Z_K]$, in which each Z_i was approximately independently and normally distributed with mean, zero, and variance, unity. The random vector $\underset{\sim}{x}_{j(i)}$ was then expressible as

$$\underset{\sim}{x}_{j(i)} = \underset{\sim}{\mu}_{(i)} + \underset{\sim\sim}{L}\underset{\sim}{z} ,$$

where $\underset{\sim}{x}_{j(i)}$ had an underlying K-variate normal distribution with mean vector $\underset{\sim}{\mu}_{(i)}$ and covariance matrix $\underset{\sim}{\Sigma}_i$.

The random numbers required to generate each vector $\underset{\sim}{z}$, and consequently each $\underset{\sim}{x}_{j(i)}$, were produced with the computer subroutine RAN3Z in the University of Illinois Computing Services Subroutine Library. This subroutine, a method of Dr. D. H. Lehmer of the University of California at Berkeley, generates a sequence of pseudo-random floating point numbers which appear to have been drawn from a population which is uniformly distributed on the closed interval 0 to 1. These pseudo-random numbers meet all the statistical requirements of truly random numbers, yet given the same starting integer, the identical sequence of pseudo-random numbers can be reproduced at any given time - consequently arises the name pseudo.

The method of transforming a sequence of uniformly distributed random numbers to one which is normally distributed, as was required for the vector $\underset{\sim}{z}$, was the Box-Muller (1958) method for the generation of normal deviates. The Box-Muller method converts a pair of random

numbers R_1 and R_2 between 0 and 1 from a uniform distribution to the random deviates Z_1 and Z_2 from a standard normal distribution by the following relations:

$$Z_1 = (-2 \log_e R_1)^{1/2} \cos(2\pi R_2),$$

and

$$Z_2 = (-2 \log_e R_1)^{1/2} \sin(2\pi R_2) .$$

2.3.4 Procedure for Obtaining Significance Level Estimates

The general procedure for obtaining significance level estimates for the conventional mixed model ANOVA, conservative ANOVA, and multivariate tests under various specifications of $\underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_I$, and $\underset{\sim}{\Sigma}$ was as follows:

1. Fix the value of $\underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_I$, and $\underset{\sim}{\Sigma}$ (for present purposes I = 3).
2. Set null values of the K-variate mean vectors $\underset{\sim}{\mu}_{(i)}$ (in the current study, let $\underset{\sim}{\mu}'_{(i)} = [0,0,0,0]$.
3. Compute the following statistics for factors A, B, and A x B for a generated random sample of observations (N = 15 in the present study):
   a. For A-effects compute $\mathfrak{I}_A$.
   b. For B-effects compute $\mathfrak{I}_B$ and $\mathcal{T}_B$.
   c. For A x B-effects compute $\mathfrak{I}_{AB}$ and Rao's $\mathcal{R}$.
4. For α = .10, .05, .025, and .01 record separately whether
   a. $\mathfrak{I}_A > \mathfrak{I}_{1-\alpha}[(I-1),(N-I)]$
   b. $\mathfrak{I}_B > \mathfrak{I}_{1-\alpha}[(K-1),(K-1)(N-I)]$
   c. $\mathfrak{I}_B > \mathfrak{I}_{1-\alpha}[1,(N-I)]$

d. $\mathcal{F}_B > \mathcal{F}_{1-\alpha}[(K-1),(N-I-K+2)]$

e. $\mathcal{F}_{AB} > \mathcal{F}_{1-\alpha}[(I-1)(K-1),(K-1)(N-I)]$

f. $\mathcal{F}_{AB} > \mathcal{F}_{1-\alpha}[(I-1),(N-I)]$

g. $\mathcal{R} > \mathcal{F}_{1-\alpha}[2(K-1),2(N-K-1)]$.

5. Repeat parts 3 and 4 for 2000 independent samples.

6. Compute the estimates of level of significance for the seven tests; i.e., the proportion of the 2000 samples which meets individually the conditions in part 4.

A FORTRAN computer program for the IBM 360 was written by the author to obtain the significance level estimates according to the preceding general procedure. The program was designed to accommodate those configurations wherein the number of occasions was ten or less ($K \leq 10$) and the number of treatment levels did not exceed ten ($I \leq 10$). No restriction was placed upon the size of the random sample of observations; i.e., N had no upper limit. The methods of Section 2.3.3 were employed to computer simulate multivariate data from the $\mathcal{N}(\underset{\sim}{0},\underset{\sim}{\Sigma}_i)$ distribution.

The coefficients in the computer program which were used to obtain the population treatment covariance matrices from the matrix $\underset{\sim}{\Sigma}$ were computed in the following manner:

Let c_i be the coefficient for the treatment level i, n_i and a_i be the sample size and ratio coefficient, respectively, for the treatment level i (as given in Table 2), and N be the sample size across all treatment levels. Then the following relationships hold:

$$r_i = n_i/N, \quad \underset{\sim}{\Sigma} = \sum_{i=1}^{I} (r_i \underset{\sim}{\Sigma}_i), \quad \underset{\sim}{\Sigma}_i = c_i \underset{\sim}{\Sigma} ,$$

where

$$c_i = \frac{a_i}{a_1 r_1 + a_2 r_2 + \ldots + a_I r_I} .$$

As an example of how the coefficients were employed, suppose that one wished to obtain the three population treatment covariance matrices in set 321 of the present study. From set 2 in Table 3 and set 1 in Table 4 we have

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 100 & 40 & 60 & 80 \\ 40 & 400 & 120 & 160 \\ 60 & 120 & 900 & 240 \\ 80 & 160 & 240 & 1600 \end{bmatrix} .$$

Set 3 in Table 2 yields the values

$a_1 = 1$, $a_2 = 1$, $a_3 = 3$, $n_1 = 5$, $n_2 = 5$, $n_3 = 5$, and $N = 15$.

Thus,

$r_1 = 1/3$, $r_2 = 1/3$, $r_3 = 1/3$, $c_1 = .6$, $c_2 = .6$, and $c_3 = 1.8$.

Finally, it is found that

$$\underset{\sim}{\Sigma}_1 = c_1\underset{\sim}{\Sigma} = \begin{bmatrix} 60 & 24 & 36 & 48 \\ 24 & 240 & 72 & 96 \\ 36 & 72 & 540 & 144 \\ 48 & 96 & 144 & 960 \end{bmatrix} = \underset{\sim}{\Sigma}_2 ,$$

and

$$\underset{\sim}{\Sigma}_3 = c_3\underset{\sim}{\Sigma} = \begin{bmatrix} 180 & 72 & 108 & 144 \\ 72 & 720 & 216 & 288 \\ 108 & 216 & 1020 & 432 \\ 144 & 288 & 432 & 2880 \end{bmatrix} .$$

CHAPTER 3

RESULTS

Empirical results on level of significance for the conventional mixed model ANOVA, Greenhouse-Geisser conservative ANOVA, and multivariate procedures for the repeated measures model are presented in this chapter. The 96 combinations of treatment sample sizes and covariance matrices for which results were computed are referred to by their unique three-digit identification numbers which were derived in the manner detailed in Section 2.3.2.

The estimated levels of significance are examined in this chapter from the following perspectives:

1. Results are presented individually for the seven tests of A, B, and A x B-effects.
2. For each of the three testing procedures, the results for relevant effects are compared.
3. The effect upon level of significance of varying treatment sample size ratios is examined for all seven tests.
4. The results of the present study are compared with those of two previous studies.
5. A summarization is given on the effects upon level of significance of violating one or both of the covariance homogeneity assumptions.

Each of the 96 sets has associated with it Box's ϵ-index (see Section 1.3) which is a measure of the extent to which the covariance matrices of each set depart from the compound symmetry pattern. The values of ϵ for these 96 sets are presented in Table 5. These values

Table 5

Values of Box's ε-index for 96 Combinations of Covariances and Treatment Sample Sizes

Sets*	ε-index
111-811	1.000
121-821	0.736
131-831	0.725
112-812	1.000
122-822	0.624
132-832	0.597
113-813	0.901
123-823	0.635
133-833	0.617
114-814	0.533
124-824	0.492
134-834	0.452

*Sets 111-811, for example, include the following sets: 111,211,311,411,511,611,711, and 811.

will be used in subsequent discussions on the effect upon level of significance of violating the compound symmetry assumption for the covariance matrix Σ.

3.1 Empirical Results on Testing the Significance of Factors A, B, and A x B

3.1.1 Test of A-effects

The estimated levels of significance for the conventional test of treatment effects are presented in Tables 6, 7, and 8 . The results for the 36 sets in which the treatment sample sizes are equal (i.e., $n_1 = n_2 = n_3 = 5$) are contained in Table 6. The data for the 36 sets in which $n_1 > n_2 > n_3$ are presented in Table 7, while Table 8 shows the results for the remaining 24 sets in which $n_1 < n_2 < n_3$. Results for all 96 sets are shown at the nominal significance levels of .10, .05, .025, and .01.

As an example of how the tables are read, let us arbitrarily choose set 631 in Table 7 . For this situation it can be said that if three treatments are compared and $n_1 = 9$, $n_2 = 5$, and $n_3 = 1$, and if the population treatment covariance matrices Σ_1, Σ_2, and Σ_3 are in the numerical ratios of 1:1:3, and if the population covariance matrix over all treatments Σ has diagonal elements $\sigma_{11} = 100$, $\sigma_{22} = 225$, $\sigma_{33} = 400$, $\sigma_{44} = 900$, and off-diagonal elements whose intercorrelations all equal .20; then the actual probabilities of Type I error are approximately .273 when the experimenter believes it is .10, .182 instead of .05, .121 rather than .025, and .067 when the probability is thought to be .01. For this set, the null distribution of $\mathfrak{F}_A$ has been shifted to the right (increasing its values) and the actual significance levels far exceed the corresponding nominal significance levels. The experimenter

Table 6

Estimated Levels of Significance Using the Conventional Test of A-effects (Treatment Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 \leq \underset{\sim}{\Sigma}_3$ and $n_1 = n_2 = n_3$

Set	ϵ	Nominal α-level: .10	.05	.025	.01
111	1.00	.104	.052	.024	.010
121	.74	.085	.036	.018	.006
131	.72	.098	.045	.020	.008
112	1.00	.098	.052	.026	.010
122	.62	.098	.050	.020	.010
132	.60	.094	.042	.020	.008
113	.90	.086	.044	.022	.011
123	.64	.112	.057	.026	.007
133	.62	.097	.058	.024	.008
114	.53	.104	.057	.024	.010
124	.49	.100	.046	.022	.010
134	.45	.098	.044	.027	.011
211	1.00	.103	.056	.028	.014
221	.74	.105	.061	.032	.012
231	.72	.106	.052	.025	.012
212	1.00	.104	.050	.028	.016
222	.62	.101	.050	.024	.011
232	.60	.110	.056	.031	.014
213	.90	.112	.066	.036	.017
223	.64	.116	.065	.034	.014
233	.62	.100	.060	.032	.013
214	.53	.104	.053	.024	.012
224	.49	.114	.066	.040	.018
234	.45	.099	.054	.028	.011
311	1.00	.113	.058	.034	.015
321	.74	.114	.058	.027	.008
331	.72	.112	.062	.036	.014
312	1.00	.105	.056	.032	.014
322	.62	.101	.057	.028	.010
332	.60	.110	.060	.031	.012
313	.90	.125	.076	.036	.020
323	.64	.112	.062	.034	.014
333	.62	.112	.054	.030	.011
314	.53	.110	.056	.034	.014
324	.49	.112	.058	.032	.014
334	.45	.096	.050	.030	.015

Table 7

Estimated Levels of Significance Using the Conventional Test of A-effects (Treatment Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$ and $n_1 > n_2 > n_3$

Set	ϵ	Nominal α-level: .10	.05	.025	.01
411	1.00	.146	.088	.048	.021
421	.74	.156	.094	.058	.031
431	.72	.179	.104	.060	.032
412	1.00	.142	.078	.044	.016
422	.62	.158	.098	.053	.029
432	.60	.152	.086	.054	.026
413	.90	.153	.086	.054	.032
423	.64	.162	.098	.061	.028
433	.62	.164	.098	.059	.026
414	.53	.165	.108	.067	.034
424	.49	.160	.098	.056	.026
434	.45	.180	.108	.062	.033
511	1.00	.168	.104	.062	.035
521	.74	.194	.114	.061	.029
531	.72	.164	.102	.068	.030
512	1.00	.183	.110	.070	.033
522	.62	.170	.098	.063	.036
532	.60	.162	.098	.064	.029
513	.90	.183	.107	.068	.029
523	.64	.176	.100	.059	.036
533	.62	.186	.106	.066	.034
514	.53	.165	.105	.064	.032
524	.49	.170	.098	.052	.029
534	.45	.172	.102	.060	.028
611	1.00	.253	.178	.119	.066
621	.74	.259	.168	.106	.054
631	.72	.273	.182	.121	.067
612	1.00	.257	.164	.112	.066
622	.62	.260	.178	.120	.066
632	.60	.253	.175	.108	.064
613	.90	.254	.170	.110	.064
623	.64	.271	.176	.122	.075
633	.62	.246	.156	.105	.056
614	.53	.263	.181	.127	.074
624	.49	.290	.194	.130	.084
634	.45	.261	.170	.115	.059

Table 8

Estimated Levels of Significance Using the Conventional Test of A-effects (Treatment Effects) for Sets in which $\Sigma_1 \leq \Sigma_2 < \Sigma_3$ and $n_1 < n_2 < n_3$

Set	ε	Nominal α-level: .10	.05	.025	.01
711	1.00	.066	.031	.016	.008
721	.74	.078	.036	.020	.010
731	.72	.066	.036	.016	.006
712	1.00	.071	.035	.016	.006
722	.62	.065	.029	.013	.004
732	.60	.064	.036	.020	.008
713	.90	.064	.026	.012	.005
723	.64	.059	.027	.012	.006
733	.62	.069	.034	.018	.006
714	.53	.055	.028	.016	.008
724	.49	.067	.034	.016	.006
734	.45	.074	.039	.020	.008
811	1.00	.030	.013	.007	.003
821	.74	.034	.014	.006	.002
831	.72	.043	.018	.008	.004
812	1.00	.037	.014	.006	.002
822	.62	.032	.014	.008	.004
832	.60	.044	.020	.008	.004
813	.90	.032	.014	.006	.002
823	.64	.036	.016	.009	.002
833	.62	.037	.013	.005	.002
814	.53	.036	.015	.006	.002
824	.49	.035	.014	.006	.002
834	.45	.038	.016	.006	.002

is in the position of having a much greater chance of rejecting a true null hypothesis than he realizes.

It can be concluded that an inspection of the data for factor A that the degree to which $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$ differed from the compound symmetry pattern had no effect whatsoever on the estimated level of significance. This can be seen by comparing the first 12 sets in Table 6. These sets have associated ε-values ranging from 1.000 (in which case the compound symmetry assumption is satisfied) to .452 (where an extreme violation of the assumption occurs). Since none of these sets violates the assumption of homogeneity of covariance matrices for treatment levels, and since none of the estimated levels of significance differs from the nominal α-levels in more than small random amounts, it is apparent that the conventional test of treatment effects is robust (insensitive) to violation of the assumption of compound symmetry.

Violation of the assumption of equality of treatment covariance matrices (i.e., that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_3$) had serious repercussions upon estimated level of significance. In cases of equal treatment sample sizes (those sets in Table 6 whose first digit is a "2" or "3"), a small overestimation of the nominal significance level usually resulted. However, in instances of unequal sample sizes, much larger disturbances occurred in the estimated level of significance. When fewer numbers of observations were sampled from treatment populations with larger covariance matrices (as was the case for all 36 sets in Table 7), disregarding the assumption shifted the distribution of $\mathfrak{F}_A$ to the right leading to overestimation of the nominal probability α. A comparison of the data for the 24 sets in Table 7 whose first digit is either a

"4" or "5" suggested that when the ε-value for a set was at or near 1.00, this overestimation is related to the extent of heterogeneity among the three treatment covariance matrices, with larger degrees of heterogeneity producing larger discrepancies in probability; however, when the ε-value was low, the relationship was much less clear. A similar comparison of the results for the 24 sets in Table 7 whose first digit is either a "5" or "6" showed that for a given N, discrepancies from probabilities were related to the treatment sample sizes, with larger discrepancies associated with larger inequalities among the treatment sample sizes. Finally, an inspection of the data for the two 12-block sets in Table 8 led to the conclusion that when greater numbers of observations were sampled from treatment populations with larger covariance matrices, violation of the assumption that $\Sigma_1 = \Sigma_2 = \Sigma_3$ caused a shifting of the distribution of $\mathfrak{F}_A$ to the left so that the probability of Type I error was less than the nominal α-level.

The preceding conclusions were similar to the one-way classification findings of Box (1954a) where moderate inequality of variance did not affect the level of significance if the groups were equal, but did cause serious discrepancies in the case of unequal sample sizes. The similarity of the findings for the two studies is not surprising because, as was pointed out in Section 1.3, the conventional test of A-effects results from a one-way ANOVA with all observations having the same variance when the Σ_1 are equal.

3.1.2 Tests of B-effects

Tables 9, 10, and 11 provide estimates of levels of significance for the conventional, conservative, and multivariate tests of occasion effects. Data are to be interpreted in the same manner as in the three tables for treatment effects.

Table 9

Estimated Levels of Significance Using Three Tests of B-effects (Occasion Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 \leq \underset{\sim}{\Sigma}_3$ and $n_1 = n_2 = n_3$

			Conventional Test				Conservative Test				Multivariate Test			
Set	ε	α:	.10	.05	.025	.01	.10	.05	.025	.01	.10	.05	.025	.01
111	1.00		.107	.052	.028	.013	.036	.010	.001	.000	.100	.050	.026	.014
121	.74		.131	.078	.046	.020	.062	.015	.002	.000	.102	.044	.018	.006
131	.72		.115	.063	.038	.022	.049	.020	.008	.004	.102	.054	.028	.012
112	1.00		.094	.045	.020	.011	.030	.008	.001	.000	.094	.045	.024	.010
122	.62		.131	.080	.048	.026	.060	.021	.008	.002	.103	.052	.028	.011
132	.60		.117	.078	.050	.031	.062	.023	.008	.002	.109	.058	.028	.012
113	.90		.105	.050	.026	.010	.034	.008	.002	.000	.105	.048	.024	.009
123	.64		.133	.084	.058	.029	.070	.023	.008	.001	.107	.052	.029	.008
133	.62		.134	.082	.053	.028	.066	.021	.008	.002	.108	.048	.020	.009
114	.53		.141	.092	.062	.037	.078	.030	.011	.002	.101	.052	.030	.012
124	.49		.124	.083	.061	.038	.070	.030	.010	.004	.098	.049	.024	.008
134	.45		.136	.100	.068	.046	.085	.039	.015	.005	.109	.054	.024	.010
211	1.00		.101	.050	.028	.010	.036	.004	.002	.000	.116	.062	.030	.013
221	.74		.128	.077	.044	.023	.056	.018	.006	.000	.118	.058	.032	.013
231	.72		.099	.058	.032	.016	.044	.012	.004	.001	.100	.054	.027	.012
212	1.00		.104	.053	.023	.010	.036	.007	.002	.000	.111	.058	.034	.014
222	.62		.125	.078	.048	.025	.061	.022	.006	.000	.116	.059	.028	.014
232	.60		.136	.090	.054	.032	.074	.027	.009	.003	.114	.062	.034	.012
213	.90		.109	.058	.034	.012	.043	.008	.002	.000	.120	.058	.030	.012
223	.64		.134	.075	.050	.030	.060	.024	.004	.000	.132	.058	.028	.012
233	.62		.118	.074	.047	.028	.058	.024	.010	.001	.108	.056	.026	.014
214	.53		.130	.089	.059	.034	.074	.028	.010	.001	.111	.057	.026	.010
224	.49		.147	.097	.064	.038	.076	.033	.015	.004	.115	.061	.030	.012
234	.45		.158	.110	.074	.046	.092	.039	.014	.005	.102	.052	.025	.012
311	1.00		.110	.066	.034	.012	.046	.008	.002	.000	.122	.068	.036	.016
321	.74		.128	.078	.044	.022	.062	.016	.008	.001	.135	.067	.035	.013
331	.72		.114	.070	.042	.021	.054	.016	.006	.001	.126	.069	.038	.016
312	1.00		.109	.060	.030	.014	.042	.010	.002	.000	.128	.070	.030	.012
322	.62		.134	.082	.053	.029	.066	.022	.008	.000	.130	.076	.040	.020
332	.60		.116	.072	.045	.025	.058	.021	.007	.001	.116	.061	.032	.014
313	.90		.117	.066	.038	.015	.048	.011	.002	.000	.130	.072	.038	.018
323	.64		.138	.086	.054	.029	.068	.024	.007	.002	.117	.063	.036	.018
333	.62		.132	.084	.054	.030	.067	.023	.008	.002	.132	.062	.034	.017
314	.53		.138	.088	.057	.032	.074	.026	.009	.002	.136	.068	.039	.014
324	.49		.145	.101	.068	.049	.082	.038	.017	.002	.120	.067	.038	.015
334	.45		.145	.098	.070	.042	.085	.039	.020	.010	.118	.064	.034	.016

Table 10

Estimated Levels of Significance Using Three Tests of B-effects (Occasion Effects) for Sets in which $\Sigma_1 \leq \Sigma_2 < \Sigma_3$ and $n_1 > n_2 > n_3$

			Conventional Test				Conservative Test				Multivariate Test			
Set	ϵ	α:	.10	.05	.025	.01	.10	.05	.025	.01	.10	.05	.025	.01
411	1.00		.118	.060	.032	.013	.044	.008	.002	.000	.124	.066	.032	.020
421	.74		.136	.080	.043	.024	.062	.020	.007	.000	.136	.070	.039	.021
431	.72		.136	.086	.054	.031	.067	.026	.007	.002	.126	.063	.035	.016
412	1.00		.112	.062	.032	.018	.044	.016	.002	.000	.126	.066	.038	.014
422	.62		.136	.090	.056	.031	.072	.025	.007	.002	.126	.070	.040	.016
432	.60		.148	.096	.065	.036	.077	.028	.008	.002	.136	.064	.031	.012
413	.90		.124	.062	.034	.016	.046	.012	.002	.000	.129	.067	.036	.017
423	.64		.131	.082	.044	.023	.068	.017	.006	.002	.124	.064	.033	.012
433	.62		.126	.082	.052	.032	.066	.024	.009	.000	.120	.052	.025	.011
414	.53		.140	.091	.060	.036	.072	.029	.010	.002	.120	.062	.033	.015
424	.49		.152	.105	.070	.049	.089	.042	.014	.004	.128	.064	.027	.012
434	.45		.154	.112	.084	.056	.098	.045	.018	.007	.122	.064	.033	.014
511	1.00		.122	.070	.039	.014	.054	.010	.002	.000	.126	.068	.036	.018
521	.74		.142	.084	.050	.028	.062	.018	.006	.001	.128	.061	.032	.012
531	.72		.125	.072	.047	.024	.058	.020	.005	.002	.116	.060	.034	.014
512	1.00		.116	.057	.034	.014	.042	.011	.002	.000	.122	.064	.036	.019
522	.62		.134	.080	.052	.030	.064	.025	.008	.002	.124	.068	.034	.015
532	.60		.130	.086	.058	.031	.074	.026	.007	.001	.128	.066	.034	.011
513	.90		.130	.075	.038	.020	.056	.014	.003	.000	.114	.065	.035	.014
523	.64		.152	.100	.062	.038	.081	.032	.011	.004	.128	.070	.034	.016
533	.62		.138	.082	.052	.030	.068	.024	.008	.003	.130	.072	.040	.016
514	.53		.137	.084	.062	.038	.070	.032	.014	.006	.130	.071	.038	.020
524	.49		.154	.104	.072	.044	.090	.037	.016	.004	.136	.076	.043	.017
534	.45		.179	.121	.088	.056	.101	.047	.023	.007	.136	.071	.038	.015
611	1.00		.126	.064	.035	.018	.046	.013	.002	.000	.119	.062	.030	.015
621	.74		.141	.088	.058	.030	.072	.022	.008	.002	.128	.069	.037	.014
631	.72		.136	.091	.058	.032	.075	.027	.010	.002	.132	.074	.034	.013
612	1.00		.115	.062	.036	.014	.048	.010	.002	.000	.132	.070	.036	.014
622	.62		.158	.101	.068	.044	.081	.034	.012	.002	.140	.066	.038	.018
632	.60		.147	.100	.066	.034	.080	.029	.012	.004	.122	.068	.035	.013
613	.90		.134	.085	.048	.022	.062	.015	.003	.000	.125	.072	.036	.016
623	.64		.163	.098	.064	.033	.076	.026	.012	.004	.136	.072	.042	.015
633	.62		.158	.106	.072	.040	.090	.030	.010	.002	.133	.070	.034	.013
614	.53		.167	.108	.068	.038	.088	.031	.015	.004	.121	.063	.036	.020
624	.49		.172	.118	.082	.048	.098	.036	.016	.005	.113	.058	.026	.010
634	.45		.174	.124	.092	.058	.104	.050	.018	.008	.128	.066	.040	.020

Table 11

Estimated Levels of Significance Using Three Tests of B-effects (Occasion Effects) for Sets in which $\Sigma_1 \leq \Sigma_2 < \Sigma_3$ and $n_1 < n_2 < n_3$

		Conventional Test				Conservative Test				Multivariate Test			
Set	ϵ α:	.10	.05	.025	.01	.10	.05	.025	.01	.10	.05	.025	.01
711	1.00	.090	.042	.018	.008	.029	.006	.002	.000	.092	.040	.019	.008
721	.74	.124	.077	.046	.022	.058	.018	.005	.001	.110	.060	.029	.011
731	.72	.105	.060	.032	.018	.042	.013	.006	.000	.092	.048	.024	.009
712	1.00	.092	.051	.026	.012	.037	.007	.001	.000	.112	.052	.028	.011
722	.62	.120	.076	.042	.022	.054	.018	.004	.002	.108	.052	.024	.012
732	.60	.112	.070	.046	.027	.056	.022	.008	.002	.095	.050	.021	.010
713	.90	.104	.056	.031	.014	.042	.010	.002	.000	.100	.046	.021	.006
723	.64	.120	.078	.054	.031	.062	.024	.006	.002	.107	.047	.023	.008
733	.62	.122	.074	.044	.020	.056	.016	.007	.002	.088	.040	.024	.012
714	.53	.130	.086	.056	.033	.066	.026	.011	.004	.094	.051	.030	.011
724	.49	.139	.095	.063	.036	.078	.029	.012	.003	.108	.051	.024	.008
734	.45	.146	.103	.076	.050	.085	.044	.016	.006	.104	.050	.027	.010
811	1.00	.104	.044	.024	.010	.033	.005	.000	.000	.103	.054	.024	.010
821	.74	.108	.062	.036	.018	.048	.014	.004	.000	.100	.050	.024	.012
831	.72	.112	.063	.036	.022	.047	.016	.005	.000	.105	.052	.028	.012
812	1.00	.092	.046	.025	.008	.034	.005	.000	.000	.096	.057	.029	.012
822	.62	.101	.064	.039	.022	.052	.018	.006	.004	.096	.053	.028	.009
832	.60	.111	.066	.040	.020	.057	.018	.008	.002	.102	.052	.022	.009
813	.90	.089	.048	.030	.014	.038	.010	.002	.000	.087	.049	.026	.010
823	.64	.126	.074	.049	.030	.060	.024	.010	.002	.114	.066	.038	.016
833	.62	.120	.076	.048	.030	.058	.023	.006	.002	.110	.058	.032	.012
814	.53	.116	.072	.048	.026	.055	.022	.008	.002	.091	.053	.026	.010
824	.49	.122	.081	.054	.030	.064	.024	.007	.002	.090	.046	.021	.006
834	.45	.128	.088	.057	.032	.071	.028	.012	.003	.108	.060	.030	.012

The conventional test was not robust to violation of the compound symmetry assumption. The results for the first 12 sets in Table 9 indicated that the estimated level of significance was larger than the nominal α-level, increasingly so as the value of ϵ decreased from the assumed value of 1.00. The relationship was not perfect, but this could well be due to some randomness in the results caused by the standard errors of obtained estimates for the levels of significance. If we had been able to increase the number of independent computer-simulated samples from 2000 (the total for this study) to a much larger number, we could then be certain that these estimates were the actual levels of significance. Despite this impossibility, it can be seen that the results for the same 12 sets at the .01 nominal α-level, in which the standard errors are smaller than at the higher nominal values (see Table 15), closely conform to the suggested relationship between ϵ-values and estimated levels of significance.

Violation of the assumption that $\Sigma_1 = \Sigma_2 = \Sigma_3$, either by itself or in conjunction with violation of the compound symmetry assumption, had a slight effect upon the empirical level of significance. Those sets in which fewer observations were sampled from treatment levels with larger population covariance matrices produced slightly larger significance level estimates than those sets in which the reverse was true, the differences between the sets being largest when ϵ was smallest.

The conservative test underestimated the nominal α-levels in all instances except for sets 534 and 634 in Table 10 (for which $\epsilon = .452$) at the .10 nominal α-level. As was expected, the estimated levels of significance for each set were lower than those estimates obtained with the conventional test. The conservative and conventional tests were

alike in that there was a tendency for the estimated significance level to increase as the value of ϵ decreased. The effect using the conservative test on the level of significance when $\underset{\sim}{\Sigma}_1 \neq \underset{\sim}{\Sigma}_2 \neq \underset{\sim}{\Sigma}_3$ and $n_1 \neq n_2 \neq n_3$ was similar to that reported for the conventional test.

The use of the multivariate test in situations where only the compound symmetry assumption was violated did not produce noticeable discrepancies between estimated values for significance level and nominal α-levels. The multivariate test, however, was not robust to violation of the assumption of equal population treatment covariance matrices. In sets where $\epsilon \geq .90$ and it was not true that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_3$, empirical results for the multivariate and conventional tests were in close agreement, while for those sets characterized by $\epsilon \leq .74$ the multivariate test produced results closer to nominal α-levels. This was not due to any improvement in the estimability of the multivariate test, but rather to a deterioration in the ability of the conventional test to ascertain the nominal level of significance as ϵ decreased in value from 1.00.

3.1.3 Test of A x B-effects

The estimated significance levels for the conventional, conservative, and multivariate tests of interaction effects appear in Tables 12, 13, and 14. In those cases where only the compound symmetry assumption was violated (10 of the first 12 sets in Table 12), the conventional test tended to overestimate the nominal level of significance, increasingly so (generally speaking) as the value of ϵ decreased from 1.00. The magnitude of these discrepancies was, on the whole, the same as those for the conventional test of occasion effects. If the covariance matrices for the treatment levels were heterogeneous, the

Table 12

Estimated Levels of Significance Using Three Tests of A x B-effects (Treatment x Occasion Interaction Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 \leq \underset{\sim}{\Sigma}_3$ and $n_1 = n_2 = n_3$

		Conventional Test				Conservative Test				Multivariate Test			
Set	ϵ	α: .10	.05	.025	.01	.10	.05	.025	.01	.10	.05	.025	.01
111	1.00	.106	.048	.022	.010	.021	.004	.001	.000	.104	.054	.027	.008
121	.74	.123	.060	.030	.014	.029	.008	.000	.000	.093	.042	.020	.006
131	.72	.134	.076	.046	.024	.045	.012	.006	.001	.106	.051	.028	.015
112	1.00	.099	.044	.024	.010	.022	.003	.001	.000	.104	.053	.022	.008
122	.62	.131	.080	.050	.030	.048	.017	.006	.001	.095	.054	.022	.012
132	.60	.134	.077	.047	.026	.046	.014	.004	.001	.096	.049	.022	.008
113	.90	.100	.046	.019	.009	.018	.006	.002	.000	.091	.046	.024	.008
123	.64	.148	.092	.062	.032	.058	.019	.006	.001	.103	.051	.020	.010
133	.62	.128	.076	.045	.028	.044	.014	.004	.000	.098	.048	.024	.009
114	.53	.140	.094	.060	.035	.058	.019	.008	.002	.096	.050	.022	.010
124	.49	.147	.093	.062	.040	.062	.023	.010	.004	.096	.049	.022	.010
134	.45	.156	.104	.072	.048	.071	.031	.014	.004	.108	.054	.030	.014
211	1.00	.108	.055	.034	.018	.034	.010	.003	.000	.121	.065	.033	.014
221	.74	.125	.074	.043	.022	.043	.013	.006	.000	.120	.061	.031	.014
231	.72	.117	.064	.036	.019	.036	.013	.006	.002	.116	.063	.030	.013
212	1.00	.120	.066	.032	.014	.032	.006	.001	.000	.116	.066	.030	.012
222	.62	.150	.086	.056	.033	.054	.018	.005	.001	.112	.059	.029	.012
232	.60	.152	.102	.058	.037	.058	.026	.008	.002	.119	.060	.031	.010
213	.90	.096	.055	.031	.015	.030	.006	.001	.000	.093	.046	.022	.011
223	.64	.140	.088	.056	.030	.054	.014	.006	.002	.121	.064	.031	.011
233	.62	.150	.098	.061	.036	.060	.024	.009	.002	.110	.060	.028	.012
214	.53	.147	.100	.071	.041	.070	.028	.012	.004	.125	.070	.036	.014
224	.49	.170	.116	.080	.054	.078	.035	.018	.004	.118	.062	.032	.012
234	.45	.160	.104	.070	.044	.068	.029	.010	.004	.108	.056	.026	.008
311	1.00	.120	.059	.032	.016	.032	.008	.000	.000	.136	.068	.040	.016
321	.74	.128	.077	.045	.023	.044	.013	.004	.001	.112	.065	.033	.016
331	.72	.138	.086	.050	.030	.050	.020	.006	.001	.124	.074	.044	.019
312	1.00	.111	.068	.036	.014	.034	.006	.002	.000	.124	.072	.037	.016
322	.62	.135	.087	.060	.035	.060	.021	.006	.003	.123	.063	.036	.018
332	.60	.136	.086	.058	.034	.056	.022	.010	.002	.112	.065	.038	.016
313	.90	.125	.072	.035	.016	.034	.010	.002	.000	.127	.072	.038	.014
323	.64	.152	.094	.063	.038	.062	.024	.010	.004	.134	.074	.044	.020
333	.62	.145	.098	.062	.039	.060	.024	.008	.002	.117	.053	.028	.013
314	.53	.156	.108	.072	.052	.071	.034	.016	.005	.132	.072	.036	.018
324	.49	.163	.109	.080	.054	.078	.038	.016	.004	.124	.070	.034	.014
334	.45	.171	.126	.092	.060	.089	.042	.020	.010	.118	.064	.034	.016

Table 13

Estimated Levels of Significance Using Three Tests of A X B-effects (Treatment X Occasion Interaction Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$ and $n_1 > n_2 > n_3$

Set	ε	α:	Conventional Test .10	.05	.025	.01	Conservative Test .10	.05	.025	.01	Multivariate Test .10	.05	.025	.01
411	1.00		.201	.120	.076	.038	.072	.020	.004	.000	.190	.104	.064	.030
421	.74		.232	.143	.089	.052	.087	.034	.014	.006	.196	.111	.062	.030
431	.72		.212	.142	.086	.056	.084	.034	.010	.004	.197	.104	.054	.024
412	1.00		.210	.126	.082	.044	.080	.025	.008	.001	.201	.116	.071	.035
422	.62		.222	.160	.114	.070	.113	.050	.023	.010	.194	.124	.074	.036
432	.60		.217	.143	.102	.067	.100	.048	.024	.006	.206	.122	.070	.036
413	.90		.208	.128	.089	.049	.086	.027	.010	.002	.190	.110	.064	.026
423	.64		.224	.155	.110	.070	.108	.043	.016	.006	.203	.126	.072	.034
433	.62		.223	.148	.100	.058	.098	.036	.014	.002	.188	.112	.062	.024
414	.53		.226	.175	.126	.080	.123	.056	.024	.008	.200	.124	.071	.036
[illegible]	.49		.232	.169	.130	.084	.126	.056	.026	.012	.203	.124	.069	.036
434	.45		.250	.182	.136	.096	.134	.072	.036	.012	.184	.113	.068	.030
511	1.00		.244	.160	.098	.054	.096	.034	.010	.001	.230	.142	.080	.036
521	.74		.231	.155	.108	.066	.104	.040	.015	.004	.202	.118	.069	.030
531	.72		.236	.165	.114	.072	.112	.042	.018	.006	.216	.122	.072	.038
512	1.00		.230	.146	.094	.058	.092	.036	.012	.002	.218	.144	.082	.038
522	.62		.232	.168	.116	.080	.114	.056	.029	.012	.216	.134	.074	.039
532	.60		.242	.178	.134	.086	.132	.059	.028	.009	.216	.135	.080	.040
513	.90		.238	.157	.104	.061	.103	.038	.010	.005	.236	.138	.080	.042
523	.64		.244	.176	.126	.084	.126	.060	.026	.010	.212	.127	.072	.036
533	.62		.238	.180	.128	.084	.126	.060	.027	.008	.230	.136	.083	.043
514	.53		.269	.203	.150	.106	.148	.076	.037	.016	.220	.142	.087	.048
524	.49		.254	.182	.140	.096	.137	.073	.036	.014	.239	.138	.086	.046
534	.45		.252	.188	.142	.108	.141	.079	.044	.021	.216	.127	.076	.038
611	1.00		.405	.298	.226	.152	.223	.101	.039	.008	.352	.244	.163	.085
621	.74		.412	.307	.230	.176	.229	.124	.061	.022	.344	.229	.152	.088
631	.72		.393	.313	.228	.158	.224	.115	.056	.020	.355	.238	.160	.088
612	1.00		.425	.314	.233	.150	.226	.106	.038	.012	.360	.241	.162	.088
622	.62		.400	.312	.242	.176	.239	.134	.068	.026	.362	.247	.161	.080
632	.60		.396	.306	.246	.176	.242	.128	.072	.034	.346	.228	.148	.084
613	.90		.421	.309	.214	.134	.212	.092	.040	.011	.336	.220	.140	.073
623	.64		.406	.316	.242	.176	.236	.135	.070	.032	.369	.250	.166	.102
633	.62		.400	.320	.244	.181	.240	.134	.070	.024	.354	.233	.159	.089
614	.53		.404	.315	.250	.186	.248	.151	.090	.042	.363	.249	.168	.102
624	.49		.398	.318	.266	.200	.262	.159	.101	.048	.382	.264	.168	.091
634	.45		.366	.302	.247	.193	.244	.147	.091	.051	.361	.238	.152	.078

Table 14

Estimated Levels of Significance Using Three Tests of A × B-effects (Treatment × Occasion Interaction Effects) for Sets in which $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$ and $n_1 < n_2 < n_3$

Set	ε	Conventional Test .10	.05	.025	.01	Conservative Test .10	.05	.025	.01	Multivariate Test .10	.05	.025	.01
711	1.00	.050	.026	.011	.004	.011	.002	.000	.000	.060	.028	.011	.003
721	.74	.082	.046	.026	.012	.023	.006	.001	.000	.061	.028	.014	.004
731	.72	.070	.038	.020	.008	.019	.005	.002	.000	.060	.028	.014	.006
712	1.00	.049	.020	.009	.004	.008	.002	.000	.000	.062	.028	.014	.004
722	.62	.078	.046	.026	.017	.026	.006	.002	.001	.059	.028	.012	.004
732	.60	.096	.061	.038	.020	.037	.012	.004	.000	.062	.027	.016	.004
713	.90	.070	.034	.018	.007	.018	.004	.000	.000	.060	.030	.012	.006
723	.64	.074	.045	.026	.017	.026	.010	.002	.000	.045	.025	.016	.006
733	.62	.080	.050	.029	.014	.029	.005	.002	.000	.064	.024	.013	.007
714	.53	.082	.048	.032	.019	.030	.012	.004	.001	.059	.026	.008	.002
724	.49	.096	.058	.035	.017	.034	.010	.004	.002	.063	.027	.010	.004
734	.45	.111	.080	.052	.034	.050	.023	.010	.003	.078	.032	.016	.004
811	1.00	.018	.009	.004	.001	.004	.001	.000	.000	.033	.016	.008	.002
821	.74	.028	.011	.006	.003	.006	.002	.000	.000	.021	.007	.004	.003
831	.72	.036	.020	.010	.008	.010	.004	.001	.000	.038	.018	.008	.002
812	1.00	.013	.005	.002	.000	.002	.000	.000	.000	.029	.012	.006	.002
822	.62	.031	.018	.009	.004	.009	.003	.002	.000	.029	.014	.006	.002
832	.60	.042	.021	.012	.005	.012	.003	.000	.000	.030	.012	.006	.001
813	.90	.020	.010	.006	.002	.006	.002	.000	.000	.022	.011	.006	.002
823	.64	.031	.018	.007	.004	.007	.002	.000	.000	.028	.010	.004	.002
833	.62	.046	.025	.015	.007	.145	.003	.000	.000	.030	.012	.005	.002
814	.53	.040	.020	.012	.006	.011	.002	.002	.000	.023	.006	.005	.000
824	.49	.052	.030	.018	.010	.018	.006	.002	.001	.018	.007	.004	.002
834	.45	.057	.034	.023	.012	.022	.005	.002	.000	.025	.010	.002	.000

conventional test levels of significance were markedly affected by the distribution of sample sizes across the treatment levels. When the treatment sample sizes were equal (Table 12), the nominal level of significance was overestimated, but the discrepancies were much less than for those sets in which the sample sizes were unequal. The effect using the conventional test on the level of significance for unequal sample sizes resembled (but with larger discrepancies) the effect found with the conventional test of treatment effects. In sets where fewer observations were sampled from treatment levels whose population covariances were large relative to other treatment levels (Table 13), the nominal level of significance was greatly overestimated. Conversely, when treatment levels whose population covariance matrices were large relative to those of other treatment levels had more observations drawn from them (Table 14), the conventional test of interaction effects seriously underestimated the nominal probability.

The conservative test of interaction effects actually overestimated the nominal level of significance for all sets in Table 13 whose first digit is a "6". If the value of ϵ was sufficiently low, those sets whose first digit is a "4" or "5" also exaggerated the nominal α-values, particularly at the .10 nominal α-level. As was found with the conservative test of occasion effects, the conservative test of interaction effects produced lower estimated significance levels than did its conventional test counterpart. When the compound symmetry assumption was ignored, the estimated level of significance for the conservative test generally followed a pattern of increasing modestly as the value of ϵ decreased. If the premise of homogeneous treatment covariance matrices was violated, the conservative test reacted in a

manner similar to the conventional test in that instances in which $n_1 > n_2 > n_3$ and $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$ produced larger estimates of level of significance than for those cases in which $n_1 < n_2 < n_2$ and $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$.

The multivariate test of interaction effects was robust to violation of the compound symmetry assumption, as was evidenced by the close agreement between estimated and nominal significance levels. On the other hand, disregarding the assumption of equal population covariance matrices for all treatment levels did have decided consequences upon estimated level of significance, the effects being totally independent of the ϵ-value for the set of matrices. Except for a few sets whose ϵ-value was at or near its upper bound, the multivariate test data for the 96 sets were at a lower estimated significance level compared to conventional test data for the same sets; however, in every set the estimated significance level for the multivariate test was at a higher level relative to the conservative test. Within a given group of eight sets characterized by the same $\underset{\sim}{\Sigma}$ matrix (i.e., a group in which the first digit varied from "1" to "8" across all eight sets, while the second and third digits were constant), the multivariate test produced estimated levels of significance larger than nominal α-levels for all sets in which $n_1 > n_2 > n_3$ and $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$; but if $n_1 < n_2 < n_3$ and $\underset{\sim}{\Sigma}_1 \leq \underset{\sim}{\Sigma}_2 < \underset{\sim}{\Sigma}_3$, the multivariate test underestimated the nominal α-levels.

3.2 Comparison of Empirical Results for Factors A, B, and A × B

A comparison of the conventional tests of A, B, and A × B-effects shows that when the compound symmetry assumption was not adhered to, only minor random differences between estimated and nominal significance levels were found with the test of treatment effects; however, noticeable

discrepancies of the same order of magnitude in the direction of overestimation of level of significance were observed for the tests of occasion and interaction effects. These discrepancies increased in a somewhat orderly manner as the value of ϵ decreased. If the assumption of equality of population treatment covariance matrices was violated, the conventional test of occasion effects, in general, slightly overestimated the nominal α-level. Under the same conditions, larger discrepancies were noted in the test of treatment effects, and still larger discrepancies resulted with the interaction effects test. The divergencies for the latter two tests were either in the direction of overestimation or underestimation of significance, depending upon the distribution of the sample sizes.

For those sets whose population covariance matrices did not conform to the compound symmetry pattern, the conservative tests of occasion and interaction effects both produced underestimates of the nominal α-level, with the size of the discrepancies tending to diminish as the value of ϵ was reduced. On the whole, the underestimation was slightly more pronounced with the test of interaction effects. Violation of either the assumption of equal treatment covariance matrices or of both covariance matrix assumptions simultaneously had a much greater impact upon the conservative test of interaction effects than upon the conservative test of occasion effects. Depending upon the distribution of sample sizes across treatment levels, the estimates for the test of interaction effects ranged from a serious underestimation to a large overestimation of the nominal level of significance.

The multivariate tests of occasion and interaction effects did not appear to be affected by violation of the compound symmetry premise,

because they both provided estimates in general close agreement to nominal α-values. Inequality of population treatment covariance matrices caused the multivariate test of occasion effects to usually overestimate slightly the nominal level of significance. However, usage of the multivariate test of interaction effects under similar circumstances could generate either greatly inflated or reduced significance level estimates, depending upon the distribution of sample sizes across the three treatment levels.

3.3 Effect of Treatment Sample Size Differences on Empirical Results

The importance of employing equal sample sizes for the treatment levels was readily apparent from an inspection of the numerical results of this study. Of the seven statistical tests used, only the three tests of occasion effects were not seriously affected by the distribution of the sample sizes across treatment levels when the population covariance matrices of the levels were not equal.

The conventional test of treatment effects, which was robust to violation of the assumption of compound symmetry, produced discrepancies in estimated significance level which were in direct proportion to the degree to which $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$ differed relative to each other. The three tests of interaction effects were even more disturbed than was the test of treatment effects for those sets in which unequal population treatment covariance matrices were paired with unequal sample sizes.

A comparison of the multivariate tests of occasion and interaction effects for those sets wherein $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$ were in the numerical ratios of 1:2:3 or 1:1:3, but whose sample sizes for the three treatment levels were equal (i.e., sets whose first digit was a "2" or "3"),

disclosed that there was virtually no difference between the two tests in regard to how accurately they estimated nominal levels of significance. However, if the sample sizes were unequal, the multivariate test of occasion effects proved to be much less biased by violation of the equal treatment covariance matrices assumption than was the test of interaction effects. Comparing the conventional tests of occasion and interaction effects led to similar findings, except that for equal sample sizes the test of interaction effects overestimated the nominal level of significance to a slightly greater extent than did the test of occasion effects. With equal sample sizes, conservative tests of occasion and interaction effects were affected to the same small degree by violation of the assumption of equal population treatment covariance matrices; but with unequal sample sizes, the conservative test of interaction effects behaved similarly to the conventional and multivariate interaction tests in that it was much more sensitive to violations than was its corresponding test of occasion effects.

A further consequence of using equal sample sizes for the treatment levels was that those tests of treatment, occasion, and interaction effects that were not robust to violation of the assumptions concerning structure of the covariance matrices yielded slight increases in their estimated levels of significance. If equal sample sizes were used, the degree to which $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$, and $\underset{\sim}{\Sigma}_3$ differed relative to each other was directly related to increases in the overestimations of nominal significance level for all seven tests; whereas the extent to which $\underset{\sim}{\Sigma}$ differed from the compound symmetry pattern, as measured by Box's ϵ, was associated with the amount that the estimated significance level was an overestimate for the conventional and conservative tests of

occasion and interaction effects. Therefore, excepting the two conservative tests, it can be expected that using equal treatment sample sizes with the tests of treatment, occasion, and interaction effects will guarantee that the probability of Type I error will be equal to or slightly greater than the nominal probability α, no matter what may be the composition of the population treatment covariance matrices.

3.4 Comparison of Empirical Results with Those of Two Previous Studies

Collier et al. (1967) presented Monte Carlo results for the conventional tests of treatment, occasion, and treatment x occasion effects (in the terminology of the present study) and the conservative tests of occasion and interaction effects for situations in which the population covariance matrix across all treatments, $\underset{\sim}{\Sigma}$, contained elements of varying heterogeneity (assuming in all situations that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_3$). Twelve of the fifteen covariance patterns utilized by Collier et al. were also employed in the current study--namely, those sets whose first digit is a "1". The basic configuration for the 12 sets was identical in both studies--a two-factor model with three levels of a non-repeated dimension, four levels of a repeated dimension, and five subjects in each level of the non-repeated dimension. The current study differed slightly from Collier et al. in that the procedure for obtaining estimates of level of significance involved the generation of 2000 independent samples of 15 observations each, whereas Collier et al. used but 1000 samples. Computer simulating greater numbers of samples reduces the standard errors of the estimated levels of significance at the four nominal α-levels, resulting in increasingly accurate estimates.

A comparison of the standard errors of obtained estimates for the two studies is made in Table 15.

Table 15

Standard Errors of Estimated Levels of Significance at Nominal α-levels

Monte Carlo Study	α-level .10	.05	.025	.01
Collier et al. (1967)	.0095	.0069	.0049	.0031
Current Study	.0067	.0049	.0035	.0022

In general, it was found that for each of the five conventional and conservative tests, the obtained estimates of significance level were in fairly close agreement for all 12 sets that were common to both studies. A comparison of the two collections of results for the conventional test of treatment effects (for which estimated and nominal significance levels should theoretically agree) indicated that, on the whole, the estimated levels of significance for the present study, which generated twice as many random samples of observations, were more nearly equal to nominal α-values than were those of the previous study. The Collier et al. article reported that the estimated levels of significance for the conventional test of interaction effects were, for the most part, slightly more discrepant from nominal α-values than were those for the conventional test of occasion effects. This discovery was at variance with the results of the present study which did not uncover any discernible difference between the discrepancies from nominal significance levels for the conventional tests of occasion and interaction effects.

As noted in Section 2.2, the eight sets of population treatment covariance matrices and sample size ratios in Table 2 were selected from among those utilized by Ito and Schull (1964) in their study of the effect on level of significance in multivariate analysis of variance (MANOVA) of violating the assumption that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_3$. Their method of measuring the effect involved approximating the distribution of Hotelling's trace criterion τ.

Because the multivariate test of interaction effects in the present study amounts to a one-way MANOVA with the number of transformed variates equaling K-1 = 3 (see Morrison (1967), p. 188), it is possible to directly compare the results of the two studies. Table 16 contains the data obtained by Ito and Schull at the .05 nominal α-level for the approximations to the level of significance of Hotelling's trace criterion τ. The eight sets of covariance matrix patterns are the same as those in Table 2. The accompanying treatment sample size ratios (r_1, r_2, and r_3) with N tending to infinity are identical to the sample size ratios in the corresponding sets of Table 2 in which N = 15.

The estimated levels of significance for the multivariate test of interaction effects in the present study were obtained using Rao's $\mathcal{F}$-approximation statistic R which, because the degrees of freedom for the hypothesis equaled 2, followed an exact $\mathcal{F}$-distribution. Since the multivariate test is unaffected by violation of the compound symmetry assumption, closer approximations to the level of significance of Rao's R-statistic for each of the eight sets in Table 16 can be derived by averaging the empirical results in Tables 12-14 across all 12 covariance patterns of $\underset{\sim}{\Sigma}$. For a given treatment covariance pattern this

Table 16

Approximate Levels of Significance of Hotelling's Trace Criterion τ and Rao's R-statistic

(Number of Variates in Ito and Schull Study = 3)
(Number of Transformed Variates in Current Study = 3)

Set	$[\Sigma_i]$ and $[r_i]^*$ (i = 1,2,3)	Hotelling's Trace [Ito and Schull (1964)]	Rao's R-statistic [Current Study]
		.05 Nominal α-level	.05 Nominal α-level
1	Σ_1, Σ_1, Σ_1 1/3, 1/3, 1/3	.050	.050
2	Σ_1, Σ_1, Σ_1 1/3, 1/3, 1/3	.055	.061
3	Σ_1, Σ_1, $3\Sigma_1$ 1/3, 1/3, 1/3	.060	.068
4	Σ_1, $2\Sigma_1$, $3\Sigma_1$ 7/15, 5/15, 3/15	.121	.116
5	Σ_1, Σ_1, $3\Sigma_1$ 7/15, 5/15, 3/15	.147	.134
6	Σ_1, Σ_1, $3\Sigma_1$ 9/15, 5/15, 1/15	.292	.240
7	Σ_1, $2\Sigma_1$, $3\Sigma_1$ 3/15, 5/15, 7/15	.022	.028
8	Σ_1, Σ_1, $3\Sigma_1$ 1/15, 5/15, 9/15	.004	.011

*$[r_i] = n_i/N$

involves averaging the multivariate test estimates of level of significance for the block of 12 sets in either Table 12, 13, or 14 whose first digit is identical to the set number of the treatment covariance pattern in Table 16. These averages at the .05 nominal α-level are presented in Table 16 alongside the results of Ito and Schull.

The difference in Table 16 between the results of the two studies can be attributed in part to the different methods used to obtain the significance level estimates. Ito and Schull's data are based upon large sample theory and are strictly appropriate only if the treatment sample sizes are so large that the within group sums of squares and products matrix $\underset{\sim}{W}$ provides exact values for the population error matrix. The results of the present study were obtained with the use of computer simulation techniques which, while not requiring a large N, are subject to inaccuracies caused by the incapability of generating a series of variates that follow precisely a normal distribution and are completely independent of one another. In addition, the unfeasibility of generating an infinite number of independent samples introduces random errors in the obtained estimates.

A second more important contributing factor for differing results is that the two test statistics do not follow identical null distributions because there is more than one degree of freedom for effect being tested, nor is it to be expected that the two statistics should be affected similarly by violation of the hypothesis of homogeneous population treatment covariance matrices. A comparison of the results in Table 16 indicates that if the treatment sample sizes are equal, Hotelling's trace criterion τ overestimates the nominal level of significance by a smaller amount than does Rao's R-statistic; however, given unequal

sample sizes and violation of equal treatment covariance matrices, the trace criterion produces larger discrepancies from nominal probabilities than Rao's R.

3.5 Summary of Effects on Level of Significance of Violating One or Both of the Population Covariance Matrix Assumptions

The effects on level of significance of violating covariance homogeneity assumptions are summarized for the following three situations:

1. Violation of the assumption that the population covariance matrix of repeated measures is the same for each of the levels of the treatment factor - All seven statistical tests were biased by violation of this assumption, with the discrepancies largest for the tests of interaction effects and decreasing for the conventional test of treatment effects and the tests of occasion effects in that order. The conventional tests of treatment and interaction effects and the multivariate test of interaction effects all overestimated the nominal level of significance when equal treatment sample sizes were used; if the sample sizes were unequal, the tests either overestimated or underestimated the probability of Type I error, depending upon the distribution of the sample sizes across the treatment levels. In general, the conventional and multivariate tests of occasion effects slightly overestimated the nominal level of significance. The conservative tests of interaction and occasion effects behaved in a fashion akin to their conventional test counterparts in regard to any depressions or inflations of the empirical level of significance induced by the distribution of treatment sample sizes. However, because both tests were intentionally designed to underestimate the significance of effects,

the estimated significance levels were well below those of the corresponding conventional tests. In all instances the estimated levels of significance for the conservative test of occasion effects remained far below the nominal α-values, and only in extreme cases (those involving set 6 in Table 2) did the estimates for the conservative test of interaction effects exceed nominal α-values.

2. Violation of the compound symmetry assumption; i.e., that the population covariance matrix over all treatments, $\underset{\sim}{\Sigma}$, has equal variances and equal covariances - The multivariate tests of occasion and interaction effects and the conventional test of treatment effects were not affected when this assumption was disregarded. The conventional tests of occasion and interaction effects both overestimated the nominal significance level, with the discrepancies of the two tests being nearly equal for a given value of ϵ. The differences between estimated levels of significance and nominal probabilities for these two conventional tests increased in a somewhat orderly manner as the value of ϵ decreased. The conservative tests of occasion and interaction effects both experienced increases in their estimated significance levels as ϵ decreased, but these estimates remained below nominal α-values, even for sets with small ϵ-values. The empirical levels for the conservative test of interaction effects fell below nominal α-levels by larger amounts than did the empirical results for the conservative test of occasion effects.

3. Simultaneous violation of both covariance assumptions - Because the multivariate tests of occasion and interaction effects and the conventional test of treatment effects were unaffected by violation of the compound symmetry assumption, simultaneous violation of both

covariance assumptions for these three tests produced results that were identical to those reported in part 1 of this section in which only the assumption of equal population treatment covariance matrices was violated. With respect to the conventional and conservative tests of occasion and interaction effects, the following generalization may be made: for a given test the simultaneous violation of both covariance assumptions produces an estimated level of significance which, as the ϵ-value for the matrix $\underset{\sim}{\Sigma}$ decreases, tends to become increasingly <u>larger</u> than the estimated level of significance for the same test when only the assumption of equal population treatment covariance matrices is violated.

CHAPTER 4

CONCLUSIONS

4.1 Practical Implications

The results of this study were related to specific configurations, and this must be borne in mind in any generalization of the conclusions. Nevertheless, some of the implications of the empirical results may be useful to the applied researcher who, in a given repeated measures experiment, is faced with the dilemma of choosing a single test procedure.

Let us suppose that an educational researcher within a large public school system is studying the effects of different methods of training (factor A) within several school classrooms (each classroom is given a different training method) in learning a complex task. Each subject is given the same number of trials to learn the task and the trials are divided into blocks of several trials each. The blocks of trials are considered the observation units, or subject profiles, and are denoted as factor B. The preceding plan may be considered as a two-factor experiment with repeated measures on factor B, both factors being fixed. The composition of $\underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_I$, and $\underset{\sim}{\Sigma}$ would not ordinarily be known; however, because this experiment is a study of change, it is plausible to assume that $\underset{\sim}{\Sigma}$ will have the simplex form (i.e., covariances between neighboring blocks of trials will tend to be relatively higher than between blocks that are farther apart). Since the experimenter is forced to use intact groups for the levels of factor A, it may well be the case that $\underset{\sim}{\Sigma}_1 \neq \underset{\sim}{\Sigma}_2 \neq \ldots \neq \underset{\sim}{\Sigma}_I$ (e.g., a school classroom where there

are many "slow" learners would exhibit a very different pattern of mastering the complex task than would a school classroom with many "fast" learners).

The data in Chapter 3 clearly indicate the importance of using equal sample sizes for the treatment levels when the educational researcher, as in the preceding situation, has reason to believe that the covariance homogeneity assumptions were violated. If the school classrooms differ in size, our experimenter has the option of discarding observations in order to equalize the sample sizes and minimize the effects of violating the covariance assumptions. Using equal treatment sample sizes insures that the consequences of covariance assumption violations will be slight increases in the actual levels of significance for the conventional test of factor A and for the conventional, conservative, and multivariate tests of factors B and A × B. Thus, if the application of one of these statistical tests fails to show significant effects at a particular nominal α-level, our experimenter can feel assured that the effects would not be significant no matter what may be the covariance structure of $\underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_I$, and $\underset{\sim}{\Sigma}$. Conversely, if the conventional or multivariate test of either factor B or factor A × B shows significance at the nominal α-level, our experimenter should apply the corresponding conservative test to the same null hypothesis. If the effects with the conservative test using equal sample sizes are also significant, he could reasonably reject the null hypothesis of no effects; however, if the conservative test is not significant at the nominal α-level, it is questionable whether any meaningful conclusion can be reached, insofar as the rejection (or acceptance) of this hypothesis is concerned.

Our experimenter would be particularly interested in testing the treatment x blocks of trials interaction effects because this amounts to seeking an answer to the question of whether the treatment population profiles are similar in the sense that the line segments of adjacent blocks of trials are parallel. Rejection of the null hypothesis of no interaction effects implies that the treatments had a differential effect within the classroom setting (e.g., that one treatment was better than others in facilitating the learning of the task). Furthermore, the significance of interaction effects renders meaningless the tests on the main effects; instead, tests on simple main effects of factors A and B would be necessary.

Testing the significance of interaction effects is probably more important to our researcher than is testing either main effect; however, the conventional, conservative, and multivariate tests of interaction effects were all shown in Chapter 3 to be extremely sensitive to violation of the covariance homogeneity assumption in regard to the population treatment levels. Since our researcher has cause to doubt both this assumption and the assumption of compound symmetry for Σ, it is very important that for the tests of interaction effects he try to use equal, or at least nearly equal, treatment sample sizes (i.e., school classroom sizes). As for a choice between the conventional and multivariate tests of interaction effects, the results of Chapter 3 suggest that our experimenter might be wise to use the multivariate test in order to minimize consequences on the probability of Type I error. With equal sample sizes, the two tests differ little in the protection they afford to violation of the assumption of equal treatment covariance

matrices; but the multivariate test, which is robust to violation of the compound symmetry assumption, is clearly superior in instances where both assumptions are violated simultaneously, as is the suspected case for our researcher. If unequal sample sizes must be used, the multivariate test is still preferable, even for those situations where only the assumption of homogeneous treatment matrices is violated.

The tests for significance of factor B (trials) are meant to determine whether learning of the task is actually taking place over the blocks of trials. Our researcher need not be as concerned about the consequences of violating the covariance assumptions using the tests of factor B as he should be with the tests of factor A x B. Still, the use of equal sample sizes is to be recommended, whenever possible. The results of Chapter 3 indicate that our experimenter would be well-advised to use the multivariate test of trials effects in preference to the conventional test, especially since he has reason to question the validity of assuming the compound symmetry pattern for $\underset{\sim}{\Sigma}$.

4.2 Areas for Future Study

The numerical values for the 96 combinations of covariances and sample size ratios were chosen so that a comparison of results could be made with two previous studies. These numbers were not intended to represent every possible population covariance structure that one might encounter in repeated measures experiments.

The Ito and Schull (1964) study showed that the effects of unequal treatment covariance matrices on level of significance for the multivariate test of interaction effects increased as K (the number of occasions) increased. The consequences of varying K on level of

significance for the other six tests of treatment, occasion, and interaction effects could be worth investigating.

Other parameters which deserve investigation include I (the number of treatment levels) and N (total sample size) which, for the current study, were set equal to 3 and 15, respectively. Empirical results obtained by Collier et al. (1967) for the conservative tests of occasion and interaction effects and the conventional tests of treatment, occasion, and interaction effects did not appear to change when the number of observations for each treatment level was increased from 5 to 15. Box (1954a) proved mathematically that effects of variance heterogeneity on level of significance for the one-way classification would persist for large sample sizes. His results suggested that these effects could differ as the number of treatment levels is increased.

The important question of the power of the tests of treatment, occasion, and interaction effects when convariance homogeneity assumptions were violated was considered outside the scope of the present study. The Davidson (1972) and Darnell (1966) repeated measures studies included power investigations on the hypothesis of no occasion effects in a single-factor experiment having repeated measures on the same subject. (There was no treatment factor.) The two writers reached different conclusions in that Davidson favored the usage of Hotelling's T^2 test to detect occasion effects in situations which involved compound symmetry violations, whereas Darnell found that the conventional test was superior. Ito and Schull (1964) examined the effects of violating the assumption of equal treatment covariance matrices on the power of Hotelling's trace criterion τ in the one-way MANOVA. They found that if the treatment sample sizes were equal, moderate inequality of the

covariance matrices did not seriously affect the power of the test, as long as the sample sizes were very large; but when unequal sample sizes were employed, quite large effects occurred. If fewer persons were sampled from populations with larger covariance matrices, the power was overestimated; if greater numbers of persons were sampled from the populations with larger covariance matrices, the power of Hotelling's trace criterion τ was less than would be expected assuming equal treatment covariance matrices.

The power curves of the seven tests in the current study of treatment, occasion, and interaction effects could be approximated at predetermined values for level of significance, treatment sample sizes, and treatment covariance patterns by using Monte Carlo simulation techniques. The FORTRAN computer program used in this study could be changed to accommodate power investigations in situations where the covariance homogeneity assumptions were violated. Of interest in such investigations would be the proportion of the 2000 computer simulated statistics for factors A, B, and A X B which exceed their respective null distribution values at the nominal α-level for various specifications of n_i, r_i, $\underset{\sim}{\Sigma}_i$, and $\underset{\sim}{\mu}_{(i)}$ $(i = 1,2,\ldots,I)$. It should be noted that since the K-variate population mean vector $\underset{\sim}{\mu}_{(i)}$ $(i = 1,2,\ldots,I)$ need no longer be the null vector $(0,0,\ldots,0)$, it will no longer be necessarily true that $\underset{\sim}{\Sigma} = \sum_{i=1}^{I} (r_i \underset{\sim}{\Sigma}_i)$.

Finally, the effects of non-normal distributions combined with covariance homogeneity violations could be investigated. Mendoza et al. (1974) obtained Monte Carlo results for the univariate and multivariate treatment by occasion designs (three treatments with nine subjects/level and four occasions) when only the compound symmetry and normality assumptions were violated. Their results indicated that the multivariate

occasion effects test produced empirical significance levels somewhat higher than nominal levels under non-normality conditions, but that the multivariate interaction and the univariate occasion and interaction effects tests appeared to be robust to non-normality. These findings lend support to the position that normality assumptions are of less importance than covariance homogeneity assumptions, particularly so with unequal sample sizes.

REFERENCES

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Bartlett, M. S. Multivariate analysis. Journal of the Royal Statistical Society, Series B, 1947, 9, 176-197.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302. (a)

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498. (b)

Box, G. E. P. and Muller, M. E. A note on the generation of normal deviates. Annals of Mathematical Statistics, 1958, 29, 610-611.

Collier, R. O., Jr. Analysis of variance for correlated observations. Psychometrika, 1958, 23, 223-236.

Collier, R. O., Jr., Baker, F. B., Mandeville, G. K., and Hayes, T. F. Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 1967, 32, 339-353.

Cooley, W. W. and Lohnes, P. R. Introduction to statistical procedures: with computer exercises. New York: Wiley, 1968

Danford, M. B., Hughes, H. M., and McNee, R. C. On the analysis of repeated-measurements experiments. Biometrics, 1960, 16, 547-565.

Darnell, W. G. An empirical comparison of tests of the hypothesis of no fixed main effects in the mixed model. Unpublished Ph.D. dissertation, Michigan State University, 1966.

Davidson, M. L. Univariate versus multivariate tests in repeated-measures experiments. Psychological Bulletin, 1972, 77, 446-452.

Geisser, S. and Greenhouse, S. W. An extension of Box's results on the use of the $\mathfrak{F}$ distribution in multivariate analysis. Annals of Mathematical Statistics, 1958, 29, 885-891.

Gordon, G. System simulation. Englewood Cliffs, N. J.: Prentice-Hall, 1969.

Greenhouse, S. W. and Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112.

Harman, H. H. Modern factor analysis. (2nd ed.). Chicago: University of Chicago Press, 1967.

Hotelling, H. The generalization of Student's ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.

Hotelling, H. A generalized T-test and measure of multivariate dispersion. Proceedings of the Second Berkeley Symposium of Mathematical Statistics and Probability, 1951, 2, 23-41.

Ito, K. and Schull, W. J. On the robustness of the T_0^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. Biometrika, 1964, 51, 71-82.

Mendoza, J. L., Toothaker, L. E., and Nicewander, W. A. A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. Multivariate Behavioral Research, 1974, 9, 165-177.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill, 1967.

Naylor, T. H., Balintfy, J. L., Burdick, P. S., and Chu, K. Computer simulation techniques. New York: Wiley, 1966.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Roy, S. N. Some aspects of multivariate analysis. New York: Wiley, 1957.

Rulon, P. J. and Brooks, W. D. On statistical tests of group differences. Ch. 2 in D. K. Whitla (ed.), Handbook of measurement and assessment in behavioral sciences. Reading, Mass.: Addison-Wesley, 1968.

Schatzoff, M. Exact distributions of Wilks' likelihood ratio criterion. Biometrika, 1966, 53, 347-358. (a)

Schatzoff, M. Sensitivity comparisons among tests of the general linear hypothesis. Journal of the American Statistical Association, 1966, 61, 415-435. (b)

Tatsuoka, M. M. Multivariate analysis: techniques for educational and psychological research. New York: Wiley, 1971.

Wilks, S. S. Certain generalizations in the analysis of variance. Biometrika, 1932, 24, 471-494.

Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.

Yates, F. The principles of orthogonality and confounding in replicated experiments. Journal of Agricultural Science, 1933, 23, 109-145.